

KW chromatography; dementia; memory; brain function.

XX Callus domesticus.

XX PN JP08245413-A.

XX PD 24-SEP-1996.

XX PF 13-MAR-1995; 95JP-0081797.

XX PR 13-MAR-1995; 95JP-0081797.

XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX DR WPI; 1996-482127/48.

XX PT New neural spine stretching agents comprising heat shock protein 90
PT - useful as memory-improving agents, brain function activators or
PT for treating dementia

XX PS Example 1; Fig 2; 5pp; Japanese.

XX CC This is the amino acid sequence of the chicken heat shock protein-90
CC which can be used as a neural spine elongation factor. The protein was
CC isolated from chicken cells by conventional chromatography and separated
CC by SDS-PAGE. Peptide fragments were isolated from the gel and were used
CC to ascertain the amino acid sequence. The protein can be used as the
CC main ingredient in compositions e.g. to treat dementia, to improve memory
CC or to activate brain functions.

XX SQ Sequence 728 AA;

Query Match 4.7%; Score 99.5; DB 17; Length 728;
Best Local Similarity 21.3%; Pred. No. 0.58;
Matches 79; Conservative 57; Mismatches 144; Indels 91; Gaps 18;

QY 6 VPSSLSAELEAKLRFGKK--KNTHSLFVFII--PENFKGCISGHGMDIAL---TEP 57
Db 390 lphlshremllqsklikvirknlvkkclelftelaeckenkkyefqfknkligiheds 449
QY 58 LTMKMSNVVYKWTCPSTNVKTENATGPEELGL---PLQRSYSEHLGYFPTDL----- 108
Db 450 qnrkkisellrryt-----sasgdemvslkdycctrmkengkhvvyitgetkdqva 499
QY 109 -FACSESLRNGLEL-----NASLSEFEKNKIS-----LHSSKEKLRE 149
Db 500 nsafverlrk-hgleviymiepidcvcgqlkefgktlsvtkglelpedeekkkqe 558
QY 150 RIKYCCQQLTLPLPVYKGRKNDAAVLEATVDYVYIREKI-SPAVMAQITEALQSNMRF 208
Db 559 ekakafnlgkim-----kdilekvekvvvshrlvtspccivstvygwtnmer 608
QY 209 CKKQQTPIELSLPGTYMAQRENSVMSTYSPERGLOFLTNTCMNGCSTPDAESSLDEAVRV 268
Db 609 imkaq-----alrdnstmgymaakkhlel-----npd--hslietliq 644
QY 269 PS-SSASENAIGD---PYKTHISSAALSIN--SLHTVRYIS--KVTPSYDATAVTNQNI 319
Db 645 kaeadkndksvkdlvallyetalissgfsledpqthanriymikiglgldeddaaeaa 704
QY 320 SIHLPSAMPVPV 330
Db 705 spavteemopl 715

Search completed: January 19, 2001, 10:43:54
Job time: 33644 sec


```

PF 22-APR-1994; 94WO-US04496.
XX
XX
PR 14-MAY-1993; 93US-0062443.
XX
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
XX
PI Canaani E, Croce C;
XX
XX
DR WPI; 1995-006818/01.
DR N-PSDB; Q75164.
XX
XX
PT New acute lymphocytic leukaemia gene prods. - used for the
PT diagnosis and treatment of leukaemias, partic. acute
PT lymphoblastic or nonlymphoblastic leukaemia
XX
XX
PS Example 1; Page 95-98; 207pp; English.
XX
XX
CC Oligonucleotide probes of at least 15 nucleotides derived from
CC the AF-4 sequence Q75164 (which encodes R66450) and able to identify
CC chromosome abnormalities within the AF-4 gene of chromosome 4 are
CC claimed. The AF-4 protein is relatively rich in serine (168) and
CC proline (118) compared to the average frequency of these amino
CC acids (7.1% and 4.6%, respectively).
XX
XX
SQ Sequence 1210 AA;

Query Match 5.2%; Score 108; DB 16; Length 1210;
Best Local Similarity 21.4%; Pred. No. 0.19;
Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TPCSNVTVKTENAT--GPEELGLPLQ-----RSYSEHLGY-----PPTDLFACSES 114
DB || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 891 tcgdpbkssastksnkhkdsipkrrvegksrssehkgsqgdtnpfpv-----ps 944
QY 115 LRNGLELNASLSEFEKKNKISLLH-----SSKEKLRRERIKYCEQLRTLPLPYVK- 166
DB ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 945 lpngns--xpgkpvxfdkqadlhmrakmkqkaelmdrvgkafkyleavlsfiec 1001
QY 167 -----GRKNDASVLEATVDVVKYI-----REKISPAVMAQITEALQSN 205
DB || || || || || || || || || || || || || || || || || || || || ||
DB 1002 giatesesqsksaysvsetvdlkifmelskfsdataptqekifavlcrcqsilnma 1051
QY 206 MRFCCKQQTPIELSLPTVMAQRENSVMSTYSPERGLQLTNTCWCNGCSTPDAESLDEA 265
DB || || || || || || || || || || || || || || || || || || || || ||
DB 1062 mfrckkd---laikysrtlnkhfesskvaqap-----spciastgcpplspmpsp 1110
QY 266 VRVPSSASANAIGDPYKTHISSAALSLSLHT--VRYYSKVTPSYD-----ATAVTNQN 318
DB || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1111 assvgqsgsagsgvgsgvaatistpvtiqmtssyvtitshvltafdlweqaualtrkn 1169

RESULT 9
Y29500
ID Y29500 standard; Protein; 96 AA.
XX
XX
AC Y29500;
XX
XX
DT 13-OCT-1999 (first entry)
XX
XX
DE Human lung tumour protein LT86-27 predicted amino acid sequence.
XX
XX
KW Human; lung tumour protein; therapy; diagnosis; lung cancer; vaccine;
KW immunotherapy; detection; inhibition.
XX
XX
OS Homo sapiens.
XX
XX
PN WO9938973-A2.
XX
XX
PD 05-AUG-1999.
XX
XX
PF 26-JAN-1999; 99WO-US01642.
XX
XX

```

```

PR 22-DEC-1998; 98US-0219245.
PR 28-JAN-1998; 98US-0015022.
PR 28-JAN-1998; 98US-0015029.
PR 18-MAR-1998; 98US-0040828.
PR 18-MAR-1998; 98US-0040831.
PR 23-JUL-1998; 98US-0122191.
PR 23-JUL-1998; 98US-0122192.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Frudakis TN, Lodes MJ, Mohamath R, Reed SG;
XX
XX
DR WPI; 1999-479187/40.
DR N-PSDB; Z07170.
XX
XX
PT Lung tumour specific polynucleotides for inhibiting the development
PT of lung cancer
XX
XX
PS Example 2; Page 99; 171pp; English.
XX
XX
CC The present invention describes lung tumour specific polynucleotides
CC and tumour antigens. Z07144 to Z07246 and Z08301 to Z08325 represent
CC specifically claimed polynucleotides, and Y29486 to Y29571 represent
CC amino acid sequences from the present invention. The lung tumour
CC specific polynucleotides and polypeptides can be used in pharmaceutical
CC compositions and vaccines to inhibit the development of lung cancer.
CC They can also be used to detect lung cancer in a patient. Probes and
CC antibodies derived from the lung tumour sequences are useful in
CC detection of lung cancer.
XX
XX
SQ Sequence 96 AA;

Query Match 5.0%; Score 105; DB 20; Length 96;
Best Local Similarity 33.0%; Pred. No. 0.005;
Matches 32; Conservative 15; Mismatches 40; Indels 10; Gaps 3;

QY 86 PEEGLPLQSYSEHLGYPTDLFACSESLRNGLELNASLSEFEKKNKISLLHSS--K 143
DB || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4 pevrsriptvrgcngsl-----sgalsccedsaqsgppkaptvaegps----sclrrnvis 55
QY 144 EKLRERIKYCCQLRTLPLPYVKGRKNDASVLEATV 180
DB || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 56 ererkrmslscrlralpqlqgrredmasvlemsv 92

RESULT 10
R33386
ID R33386 standard; Protein; 160 AA.
XX
XX
AC R33386;
XX
XX
DT 15-JUL-1993 (first entry)
XX
XX
DE Max protein.
XX
XX
KW Mad; max; myc; c-myc; helix-loop-helix zipper; leucine zipper;
KW helix-turn-helix; diagnosis; prognosis; cancer; malignancy;
KW neoplasm; tumour; studying embryogenesis; study gene regulation.
XX
XX
FH Key Location/Qualifiers
FT Region 9..17
FT /note= "9 amino acid insertion found in several
FT PCR clones."
FT Domain 39..50
FT /label= Helix I
FT /note= "Helix I of b-HLH homology region."
FT Domain 57..75
FT /label= Helix II
FT /note= "Helix II of b-HLH homology region"
XX
XX
PF WO9305056-A.
XX
XX

```

PD 24-NOV-1994:

mSin nucleic acids encoding recombinant polypeptide(s) that associate with Mad polypeptide - are possible homologues of Saccharomyces cerevisiae general repressor protein

Example 2; Fig 2D; 111pp; English.

PA (HUBE/) HUBERT R S.
 PA (RAIT/) RAITANO A B.
 XX Afar DE, Hubert RS, Raitano AB;
 XX WPI; 2000-237872/20.
 DR N-PSDB; 294275.
 XX
 PT Testis specific Helix Loop Helix proteins expressed in cancers and
 PT useful for the prevention, diagnosis and treatment of prostate, bladder
 PT and ovarian tumors -
 PS
 PS Claim 1; Fig 2A-D; 62pp; English.
 XX
 CC This sequence is that of human PHELIX, a novel basic Helix Loop
 CC Helix protein thought to act as a transcription factor. PHELIX
 CC normally exhibits a testis-specific expression pattern but is
 CC up-regulated in prostate and other types of cancer. The invention
 CC provides diagnostic and therapeutic methods useful in the
 CC management of various cancers which express PHELIX, including
 CC prostate cancer, bladder cancer, ovarian cancer and testicular
 CC cancer, including therapies aimed at inhibition the transcription,
 CC translation, processing or function of PHELIX. The expression
 CC pattern of PHELIX suggests that is an ideal target for a cancer
 CC vaccine approach to prostate cancer. PHELIX protein can also be
 CC used to screen for agonists and antagonists of therapeutic value
 CC and to raise antibodies.
 XX
 SQ Sequence 405 AA;

Query Match 100.0%; Score 2095; DB 21; Length 405;
 Best Local Similarity 100.0%; Pred. No. 6 4e-203;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVLKVPSSLSAELEAKLIRFGKKNTSHLSFVILPENFKGICSGHGMIALTEPLTM 60
 Db 1 mvlkvpsslsaeleaklirfgkknthslsfviiipenfkgicsgghgmialtepltm 60
 Qy 61 EKMSNVVYKWTCTSNVKTENATGPEELGLPLQRSYSEHLGYFTDLFACSESLRNG 120
 Db 61 ekmsnvvykwtctpsntvktentatgpeelglplqrsysehlgyftdlfacseslrng 120
 Qy 121 LELNASLSEFEKKNKISLLSHSSEKLRERIKYCCQLRTLPLPVYKGNDAASVLEATV 180
 Db 121 lelnaslsefeknkksllshsseklererikyccqlrtlplpyvkrkndaasvleatv 180
 Qy 181 DYVKYIREKISPVAWQITEALQSNMRFCKKQOQTPIELSLPGTYMAQRENSVMSTYSPE 240
 Db 181 dyvkyirekispvawqitealqsnmrfckkqqtptielslpgtyvmagrensvmstysper 240
 Qy 241 GLOFLTNTCKNGCSTPAEESLDEAVRPSSASENAGDPYKTHISSAALSLSLHTR 300
 Db 241 glqflntcngcstpaeesldeavrpssasenaigdpkthissaalslnslhtr 300
 Qy 301 YYSKVTSPYDATAYTNONISIHLPSPAMPVPVSSFSGLGTALLGWARRALHIPTVCNSFGRIK 360
 Db 301 yyskvtspydataytnqnishlpampvpvssfselgtallgwarrahliptvcnsfgrik 360
 Qy 361 STCKFLTSTTYWAFQNLGKVEORMILKAPPKDLISKELAWFGF 405
 Db 361 stckflsttywafqnlgkveormilkappkdliskelawfgf 405

RESULT 2
 Y67501
 ID Y67501 standard; Protein; 158 AA.
 XX Y67501;
 XX
 XX 19-MAY-2000 (first entry)
 DT
 XX PBR-associated protein (PAP)8.
 DE

XX
 KW Peripheral-type benzodiazepine receptor; PBR; PBR- associated protein;
 KW PAP; cell proliferation; cancer; cell death; cytostatic; neuroprotective;
 KW immunomodulator; antinfertility; cerebroprotective; atherosclerosis;
 KW Niemann-Pick C; tumour; Alzheimer's disease; developmental disorder;
 KW cholesterol; multiple sclerosis; stress; neurodegenerative disorder;
 KW immune disorder; stroke; PAP8.
 XX
 OS Mus sp.
 XX
 XX WO200009549-A2.
 XX
 XX 24-FEB-2000.
 XX
 XX 11-AUG-1999; 99WO-US18507.
 XX
 XX 11-AUG-1998; 98US-0096048.
 XX
 XX (GEOU) UNIV GEORGETOWN MEDICAL CENT.
 XX
 XX Papadopoulos V, Li H;
 XX WPI; 2000-224278/19.
 DR N-PSDB; 257039.
 DR
 XX Novel peripheral-type benzodiazepine receptor associated proteins used
 PT for the regulation of the peripheral-type benzodiazepine receptor -
 PT
 XX
 XX Claim 20; Page 70; 71pp; English.
 PS
 XX The invention provides isolated peripheral-type benzodiazepine receptor
 CC (PBR) associated proteins (PAPs) and nucleic acids encoding the PAP
 CC proteins. The PAP polynucleotides are a source of primers and probes for
 CC detection, isolation and amplification. PAP ligands or substrates or
 CC antibodies can be labeled and used to detect PAPs, in the diagnosis and
 CC prognosis of disease associated with increased cell proliferation, such
 CC as cancer, or reduced cell death. The diagnostic methods of the
 CC invention can be predictive of diseases involving PBR including
 CC gallstones, atherosclerosis, Niemann-Pick C, Sitosterolemia, Dystrophy,
 CC tumor proliferation, Schnyder's corneal crystalline dystrophy, brain
 CC disorders including Alzheimer's disease, cholesterol metabolism,
 CC Tellurium toxicity, Smith-Lemli-Opitz syndrome, myelinization,
 CC developmental abnormalities, demyelination, Charcot-Marie tooth
 CC disease, Pelizaeus-Merzbacher disease, Multiple sclerosis, and SLA. The
 CC methods may also be useful in prophylactic treatments, or in screening
 CC for compounds effective in prophylactic treatment. The PAPs may be
 CC used to identify inhibitors or activators which allows the identification
 CC of drugs or agents which modulate PBR activity. Inhibitors of PAP may be
 CC used in the treatment or amelioration of conditions such as stress and
 CC stroke, cancer, neurodegenerative disorders, developmental disorders,
 CC infertility and immune disorders. The present sequence represents
 CC a PAP8 polypeptide.
 XX
 SQ Sequence 158 AA;

Query Match 5.5%; Score 115; DB 21; Length 158;
 Best Local Similarity 27.7%; Pred. No. 0.0011;
 Matches 28; Conservative 24; Mismatches 41; Indels 8; Gaps 3;

Qy 140 HSSKEKLRRERIKYCCQLRTLPLPVYKGNDAASVLEATVYVKYIREKISPVAWQIT 199
 Db 63 hnrmerdrirriccdelnllvpfca-etdkattlqwtatfkyiqerhgdslkfe 121
 Qy 200 EALQSNMRFCKKQOQTPIELSLP-GTYMAQRENSVMSTYSPE 239
 Db 122 sv-----fcgktrrlkltrpeslvtcpagsglgspace 156

RESULT 3
 Y07085
 ID Y07085 standard; Protein; 732 AA.
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 01:23:10 ; Search time 69.34 Seconds
(without alignments)
199.719 Million cell updates/sec

Title: US-09-389-000-2
Perfect score: 2095
Sequence: 1 MVLKVPSSLSAELEIAKL.....MILKAPPKDLISKEAWGFG 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*

1:	/SID56/gcgdata/geneseq/geneseq/AA1980.DAT:*
2:	/SID56/gcgdata/geneseq/geneseq/AA1981.DAT:*
3:	/SID56/gcgdata/geneseq/geneseq/AA1982.DAT:*
4:	/SID56/gcgdata/geneseq/geneseq/AA1983.DAT:*
5:	/SID56/gcgdata/geneseq/geneseq/AA1984.DAT:*
6:	/SID56/gcgdata/geneseq/geneseq/AA1985.DAT:*
7:	/SID56/gcgdata/geneseq/geneseq/AA1986.DAT:*
8:	/SID56/gcgdata/geneseq/geneseq/AA1987.DAT:*
9:	/SID56/gcgdata/geneseq/geneseq/AA1988.DAT:*
10:	/SID56/gcgdata/geneseq/geneseq/AA1989.DAT:*
11:	/SID56/gcgdata/geneseq/geneseq/AA1990.DAT:*
12:	/SID56/gcgdata/geneseq/geneseq/AA1991.DAT:*
13:	/SID56/gcgdata/geneseq/geneseq/AA1992.DAT:*
14:	/SID56/gcgdata/geneseq/geneseq/AA1993.DAT:*
15:	/SID56/gcgdata/geneseq/geneseq/AA1994.DAT:*
16:	/SID56/gcgdata/geneseq/geneseq/AA1995.DAT:*
17:	/SID56/gcgdata/geneseq/geneseq/AA1996.DAT:*
18:	/SID56/gcgdata/geneseq/geneseq/AA1997.DAT:*
19:	/SID56/gcgdata/geneseq/geneseq/AA1998.DAT:*
20:	/SID56/gcgdata/geneseq/geneseq/AA1999.DAT:*
21:	/SID56/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2095	100.0	405	21	Y79269 Human testis-speci
2	115	5.5	158	21	Y67501 PBR-associated pro
3	113.5	5.4	732	20	Y07085 Renal cancer assoc
4	108.5	5.2	383	15	R62656 Petunia Ph6 gene p
5	108	5.2	151	18	W10042 Max 14. Homo sapi
6	108	5.2	160	18	W10041 Max 11. Homo sapi
7	108	5.2	1187	16	R66451 AF-4 protein (enco
8	108	5.2	1210	16	R66450 AF-4 protein (enco
9	105	5.0	96	20	Y29500 Human lung tumour
10	103	4.9	160	14	R33386 Max protein. R33
11	103	4.9	2627	19	W61347 Human telomerase R
12	100.5	4.8	241	13	R21261 VHD1.3-HUCH1 regio

13	100	4.8	861	18	W22232 Human origin of re
14	100	4.8	861	18	W14138 Human origin of re
15	99.5	4.7	728	17	W09381 Chicken heat shock
16	97.5	4.7	534	20	W82799 Human IRS-p53h pro
17	97.5	4.7	1289	20	Y27163 Peptide Seq ID No:
18	97.5	4.7	1289	21	Y56781 Human Trad protein
19	97.5	4.7	1704	15	R49657 Sequence of Heliot
20	96	4.6	412	18	W08140 Human cytokine res
21	96	4.6	436	19	W69509 Mouse Pax6 protein
22	96	4.6	436	19	W69513 Mouse Pax6 protein
23	96	4.6	436	20	Y32831 Pax6 protein seque
24	96	4.6	945	20	Y35612 C. pneumoniae prot
25	95	4.5	363	21	W90763 A. aeolicus AAEOL3
26	95	4.5	363	21	Y52035 A. aeolicus AAEOL3
27	95	4.5	363	21	Y51664 A. aeolicus AAEOL3
28	94	4.5	323	20	Y06112 Nelson Bay virus s
29	94	4.5	412	20	W99060 Human basic helix-
30	92.5	4.4	1254	11	R07503 Meroprote apical-e
31	92.5	4.4	1254	18	W24375 Diacylglycerol k1
32	92	4.4	734	13	R20210 Murine axin. Mus
33	92	4.4	992	20	W96265 Human telomerase.
34	92	4.4	2625	19	W55887 Max protein. R33
35	91	4.3	221	14	R33387 Human Mad-1. Homo
36	91	4.3	221	18	W10043 Ras p21 interactin
37	91	4.3	768	17	W00811 Human RGL protein.
38	91	4.3	768	20	Y31244 RPTP-beta amino ac
39	91	4.3	2308	15	R52580 Human RPTP-beta.
40	91	4.3	2308	15	R57902 Human prostate tum
41	90.5	4.3	551	20	Y73938 Human CDC28-#3 RNA
42	90.5	4.3	1041	20	W30613 Amino acid sequenc
43	90.5	4.3	1142	21	Y43876 Salmonella typhimu
44	89.5	4.3	425	20	Y24069 Predicted retinobl
45	89.5	4.3	928	11	R06289

ALIGNMENTS

RESULT 1

Y79269

ID Y79269 standard; Protein; 405 AA.

XX

AC Y79269;

XX

DT 03-JUL-2000 (first entry)

XX

DE Human testis-specific transcription factor PHELIX.

XX

KW PHELIX: human; testis-specific; transcription factor;

KW prostate cancer; bladder cancer; ovary cancer; testicular cancer;

KW therapy; diagnosis; vaccine.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 134..150

FT Peptide /note= "nuclear localization signal"

FT Peptide 163..169

FT Peptide /note= "nuclear localization signal"

FT Domain 140..189

FT /note= "basic Helix-Loop-Helix domain"

XX

WO200012709-A2.

XX

09-MAR-2000.

XX

31-AUG-1999; 99WO-US20137.

XX

31-AUG-1998; 98US-0098610.

PR 31-OCT-1998; 98US-0106524.

XX

(UROC*) UROGENESYS INC.

PA (AFAR) AFAR D E.

Db 454 QNRKLSLRLRYT-----SASGDEMYSLKDYCTRMKNCKHIYITGTRQDVA 503
Qy 109 -FACSESLNGLNEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
Db 504 NSAFERLKR-HGLEIYVIEPIDECYQOLKEFEKGLTVSYTKESGLELPEDEEEKKOE 562
Qy 150 RIKYCEOLRTLLPYKGRKNDASVLEATVDYVKYIRKRI--SPAUMAITEALSNMF 208
Db 563 EKKTYFENLCKTM-----KDLLEKKVEKVVVSNRLVSPCLVSTYGTWATNMR 612
Qy 209 CCKOQPTLELSLPGVMAORENSVMSTYSPERGLOFLNTCWNCGSTPDASSLDEAVY 268
Db 613 IKKAQ-----ALRDNSTGMYMAKKHLEI-----NPD--HSIETTLQ 648
Qy 269 PS-SASENAIGD----PKTHISSALSLN--SLHTVYYS--KTPSYDATATYTNONI 319
Db 649 KAEAKNKNKSVADLVLYLLETALSLSGFSLDPQTHANRYRMKILGIDDEDPTADPT 708
Qy 320 SIHLPSAMPV 330
Db 709 SAAVTEEMPL 719

RESULT 2

US-08-049-282B-2
Sequence 2, Application US/08049282B
Patent No. 5534660
GENERAL INFORMATION:
APPLICANT: CHUCK, George S.
APPLICANT: DOONER, Hugo K.
APPLICANT: COURTNEY-GUTTERSON, Neal
APPLICANT: KELLER, Janis
APPLICANT: NIJJAR, Charanjit S.
APPLICANT: RALSTON, Edward J.
TITLE OF INVENTION: PH GENES AND THEIR USES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Hourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/049,282B
FILING DATE: 16-APR-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 12176-34
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-049-282B-2

Query Match 5.2%; Score 108.5; DB 1; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01; Mismatches 72; Conservative 46; Indels 75; Gaps 15;
Matches 6 VPSLSASELEAIKILIRGKKRNTH-SLFVLIIPENFK-----GCISGHGMDIA 53

Db 63 LFGPSYDEL-----SQEDHYISQTVSTILEHLSNOSKSPSTIMGCI----- 106
Qy 54 LPEPLTMRKMSNVKYWTCPSTNTKVENATGPEELGLPLQRSYSEHLGYPTDLFACSE 113
Db 107 -----QTQSAFTRWPS-PSTVVSPLDQGANSGQMLKSLFVSVPFLHTKYQTAAE 158
Qy 114 -SLRNGLELNAS-LSEFEKNKI-----SLHSSKEKLREKRIKCYCEOLRTLLPYK 166
Db 159 VSPKSRDATYVDSSTASFRKGCSTIOEPPSGNHYLAERREKIKNERFIILRSLVPEV- 217
Qy 167 GRKNDASVLEATVDYVKYIREKISP-AVMAOITFALSNMRFCCKOQPTLELSLPGVMA 225
Db 218 -TKMDKASILGDTIYVQOLKRVQDLFARANQTA-----TLQTKDTGTVK 263
Qy 226 AQRENSVMSTYSPERGLOFLNTCWN-----CSTPDASSLDEAVRVPSSASENAIGD 280
Db 264 VLQGRG-----KRRMKIVGSGVGGQAKITASSPTTHE-ELIVQVEVSIIESDALVE 315
Qy 281 ---PYK 283
Db 316 LRCPYK 321

RESULT 3

US-08-537-715-2
Sequence 2, Application US/08537715
Patent No. 5810627
GENERAL INFORMATION:
APPLICANT: Chuck, George S.
APPLICANT: Dooner, Hugo K.
APPLICANT: Courtney-Guterson, Neal
APPLICANT: Keller, Janis
APPLICANT: Nijjar, Charanjit S.
APPLICANT: Ralston, Edward J.
TITLE OF INVENTION: PH Genes and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,715
FILING DATE: 16-FEB-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/04173
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 012176-003410US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0200
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-715-2

Query Match 5.2%; Score 108.5; DB 2; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEIAKILIRGKKKNTK-SLFFVFIIPENFK-----GCISGHGMDIA 53
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 54 LTFPLTMKMSNVVKKYWTCPSPNTVKTENATGPBELLPLORSYSEHLGIFPTDLFACSE 113
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 107 -----QTGSAFTRWPS-PSTTVSSPFLDGGATSGQWLKSLIFSVPFLHTKYQTAAE 158
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 114 -SLRNGNLELNAS-LSFEFKNKI-----SLHSSKREKLRRERIKYCCEDLRTLLPYVK 166
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 159 VSPKRDATYDSSSTASFRKGCSTIOEPPSGNHVLAERRRERKLNERTILRLSLVPEV- 217
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 167 GRKNDASVLEATVDYVYIREKISP-AVMAQITEALQSNRPFCKQOTPIELSLPGTVM 225
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 218 -TKMDKASILGDTIYVQLRKKVODLEARANOEA-----TLQTKDTGTVK 263
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 226 AORENSVSTSPERGLOFLTNCWNG-----CSTPDAESSIDEAVRPSSASASNAIGD 280
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 264 VLOGRG-----KRRMKIVEGSGVGGQAKITASSPSTTHE-ELIYQVEVSIIESDALVE 315
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 281 ---PYK 283
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 316 LRCPPK 321
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

RESULT 4
PCT-US94-04173-2
Sequence 2, Application PC/TUS9404173
GENERAL INFORMATION:
APPLICANT: PH GENES AND THEIR USES
TITLE OF INVENTION: 4
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04173
FILING DATE: 15-APR-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 12176-34-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-04173-2

Query Match 5.2%; Score 108.5; DB 4; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEIAKILIRGKKKNTK-SLFFVFIIPENFK-----GCISGHGMDIA 53
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 54 LTFPLTMKMSNVVKKYWTCPSPNTVKTENATGPBELLPLORSYSEHLGIFPTDLFACSE 113
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 107 -----QTGSAFTRWPS-PSTTVSSPFLDGGATSGQWLKSLIFSVPFLHTKYQTAAE 158
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 114 -SLRNGNLELNAS-LSFEFKNKI-----SLHSSKREKLRRERIKYCCEDLRTLLPYVK 166
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 159 VSPKRDATYDSSSTASFRKGCSTIOEPPSGNHVLAERRRERKLNERTILRLSLVPEV- 217
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 167 GRKNDASVLEATVDYVYIREKISP-AVMAQITEALQSNRPFCKQOTPIELSLPGTVM 225
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 218 -TKMDKASILGDTIYVQLRKKVODLEARANOEA-----TLQTKDTGTVK 263
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 226 AORENSVSTSPERGLOFLTNCWNG-----CSTPDAESSIDEAVRPSSASASNAIGD 280
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 264 VLOGRG-----KRRMKIVEGSGVGGQAKITASSPSTTHE-ELIYQVEVSIIESDALVE 315
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 281 ---PYK 283
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

QY 316 LRCPPK 321
DB LPPGSPYDEL-----SQEDTHYSQTVSTITLHLSNOSKFSSTIMGICIS-----106

RESULT 5
US-08-537-715-4
Sequence 4, Application US/08537715
Patent No. 5910627
GENERAL INFORMATION:
APPLICANT: Chuck, George S.
APPLICANT: Dooner, Hugo K.
APPLICANT: Courtney-Guterson, Neal
APPLICANT: Keller, Janis
APPLICANT: Nijjar, Charanjit S.
APPLICANT: Ralston, Edward J.
TITLE OF INVENTION: PH Genes and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,715
FILING DATE: 16-FEB-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 012176-003410US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-715-4

Query Match 5.2%; Score 108.5; DB 2; Length 383;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEAIKILRFCKKNTN-SLFFVFIIPENK-----GCISGHGMDIA 53
DB 64 LPGSPSYDEL-----SQEDTHYSQIVSTILEHLNSQSKFSSTIMCIS----- 107

QY 54 LTPLELMKSNVYKWTQPSNTVKTENATGDELGPIQRSYSEHLGFPIDFACSE 113
DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLARKYQTAAE 199

QY 114 -SLRNGGLELNAS-LSFEKKNKI-----SLHSKSKLRERIKYCCQQLRTLLPYK 166
DB 160 VSPKSDATTVDSSTASRRFKGCSITQEEPSGNHVALERRRREKINERFIILSLVPEV- 218

QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITEALQSNMRFCKKQCPPIELSLPGTV 225
DB 219 -TKMDKASITIGDTEYVKQLRKKVODLEARNQTEA-----TLQTKDTGYK 264

QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRVPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSSTHE-BEIVQEVSIIESDALVE 316

QY 281 ---PYK 283
DB 317 LRCPYK 322

RESULT 6
PCT-US94-04173-4
; Sequence 4, Application PC/TUS9404173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PH GENES AND THEIR USES
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04173
; FILING DATE: 15-APR-1994
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,282
; FILING DATE: 16-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 12176-34-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-04173-4

Query Match 5.2%; Score 108.5; DB 4; Length 383;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEAIKILRFCKKNTN-SLFFVFIIPENK-----GCISGHGMDIA 53
DB 64 LPGSPSYDEL-----SQEDTHYSQIVSTILEHLNSQSKFSSTIMCIS----- 107

QY 54 LTPLELMKSNVYKWTQPSNTVKTENATGDELGPIQRSYSEHLGFPIDFACSE 113
DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLARKYQTAAE 199

QY 114 -SLRNGGLELNAS-LSFEKKNKI-----SLHSKSKLRERIKYCCQQLRTLLPYK 166
DB 160 VSPKSDATTVDSSTASRRFKGCSITQEEPSGNHVALERRRREKINERFIILSLVPEV- 218

QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITEALQSNMRFCKKQCPPIELSLPGTV 225
DB 219 -TKMDKASITIGDTEYVKQLRKKVODLEARNQTEA-----TLQTKDTGYK 264

QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRVPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSSTHE-BEIVQEVSIIESDALVE 316

QY 281 ---PYK 283
DB 317 LRCPYK 322

DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLARKYQTAAE 159
QY 114 -SLRNGGLELNAS-LSFEKKNKI-----SLHSKSKLRERIKYCCQQLRTLLPYK 166
DB 160 VSPKSDATTVDSSTASRRFKGCSITQEEPSGNHVALERRRREKINERFIILSLVPEV- 218

QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITEALQSNMRFCKKQCPPIELSLPGTV 225
DB 219 -TKMDKASITIGDTEYVKQLRKKVODLEARNQTEA-----TLQTKDTGYK 264

QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRVPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSSTHE-BEIVQEVSIIESDALVE 316

QY 281 ---PYK 283
DB 317 LRCPYK 322

RESULT 7
US-07-903-710-4
; Sequence 4, Application US/07903710
; Patent No. 5302519
; GENERAL INFORMATION:
; APPLICANT: Eyer, D.E.; Eisenman, R.N.; Blackwood, E.M.; Averb, D.M.
; TITLE OF INVENTION: MAX: A HELIX-LOOP-HELIX ZIPPER PROTEIN THAT FORMS A
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
; OPERATING SYSTEM: IBM PC/386 compatible
; SOFTWARE: Word for Windows-1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,710
; FILING DATE: 19920623
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/07/756,195
; FILING DATE: 09/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Sundemo, John, S.
; REGISTRATION NUMBER: 34,446
; REFERENCE/DOCKET NUMBER: FHC-1-6550
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0727 (direct)
; TELEFAX: 1-206-224-0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; DESCRIPTION: Human helix-loop-helix zipper protein; amino acid sequence
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens; Eukaryota; Animalia; Metazoa;
; ORGANISM: Chordata; Vertebrata; Mammalia; Theria; Eutheria;
; ORGANISM: Primates; Haplorhini; Catarrhini; Hominoidea;
; IMMEDIATE SOURCE: Human lymphoid B cell Manca cell line
; US-07-903-710-4

Query Match 5.2%; Score 108; DB 1; Length 151;
Best Local Similarity 24.4%; Pred. No. 0.0023;

0Y 140 HSSKKTLRERKRYCCGCELTLLTPPVKRRKNDADSAEATVDYKAIYEIKESIPVMAQIT 159
 19 HNALEKRRDRHDKDSFHSRLRDSVPSTLQSEKKSRAQLIDKATEIYIÖYMRK-----N 69
 0Y 200 EALQSNMRCECKQÖTPIELSLPGTVMAORENSVMSTYSPERGLOPLIN-----TCMNGC 253

[illegible]


```

Db 979 GIATESQSASVSVSEVVDLIKFIIMSLKSPSDATAPTOEKIFAVLCMRCSILNMA 1038
QY 206 MRFCKOQTPLESLPGTVMAQRENSVNSTYSPERGLOFLTNTCWNCGSTPDASSIDEA 265
Db 1039 MFRCKD---IAIKYSRTLKHFESSKVAQAP-----SPCIASGTGTPSPPLSPMPSP 1087
QY 266 VRPSSSASENADDPYKTHISSAALSLSLHT--VRYKYVTPSYD---ATAVTNON 318
Db 1088 ASSVGSOSASGSSGVAAATISTPVTIQNMNTSSVYITTSVLTAFDLMEQAEALTRKN 1146

```

Search completed: January 19, 2001, 10:44:57
 Job time: 32491 sec

THIS PAGE BLANK (USPTO)

Align seg 1/1 to: AK000456 from: 1 to: 2147

```
1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGluAl 17
|||||
212 ATGTTCTCTTGAAGGTGCTCTTCACTAAGTCCGAGAGAGCTGGAAGC 261
17 ailelyslleuileargpheglyllyslyslysasnthrhisserleuphev 34
|||||
262 CATCAAGTTAATTAGATTGGCAAAAAGAAAATACACATTCACCTGTTTG 311
34 alpHeilelleProGluAsnPhelYsgLYCYsileSerGlyHISGlyMet 50
|||||
312 TTTTAAATACCTCGAAATTTAAAGGTGTATTTCAGGGCATGGATG 361
51 AspIleAlaLeuThrGluProLeuThrMetGluLysMetSerAsnValVa 67
|||||
362 GATATTGCTTAAGTACAGACACATGCAATGCAAAATGAGTATGCGT 411
67 LysTYrTPThrThrCysProSerAsnThrValLysThrGluAsnAlaT 84
|||||
412 AAAATACTGGACAACATGTCCTCAACACTGTTAAGACTGAAAAGCACA 461
84 hrcGlyProGluGluLeuGlyLeuProLeuGlnArgSerTYrSerGluHIS 100
|||||
462 CTGGGCTGAGAACTTGGATTGCCCCCTGCAGAGCTCTACAGCGAAAC 511
101 LeuGlyTYrPheProThrAspLeuPhaAlaCysSerGluSerLeuArgas 117
|||||
512 CTGGGATATTCTTCTACTGATCTATTGCTGCTGCTGAACTTTAAGGAA 561
117 ngLYAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPhGluLysAsnL 134
|||||
562 TGGCATGGGCTTGAAATTAAATCTTCGTTGTCAGAGTTGGAAGAAAACA 611
134 yslYsIleSerLeuLeuHISerSerLYsgLYLysLeuArgArgGluArg 150
|||||
612 AAAAGATCTCTCTTCTTCAATTCAGCAAGAAAACCTAAGAGGAAAGA 661
151 IleLYsTYrCYsgLYGluGluLeuArgThrLeuLeuProTYrValLYsgL 167
|||||
662 ATCAAAATATTGCTGTGAGCAGCTGCTACTCTCTCCGATGTAAAGG 711
167 YArgLYsAsnAspAlaAlaSerValLeuGluAlaThrValAspTYrValL 184
|||||
712 GAGAAAGATGATCGCGCTTCACTTCTTGAGGCAACAGTTGATTGCTGA 761
184 ystTYrIleArgGluLYsIleSerProAlaValMetAlaGlnIleThrGlu 200
|||||
762 AATATATCCGGAGAAAATCTCTCCAGCGTTATGGCCCCAGATTACAGA 811
201 AlaLeuGlnSerAsnMetArgPheCYsLYsgLYGluGlnThrProIleG 217
|||||
812 GCACCTCAAGACAACATGAGGTTTGTAAAGAAAACAACACCCATTTGA 861
217 uleuSerLeuProGlyTYrValMetAlaGlnArgGluAsnSerValMetS 234
|||||
862 GCTGTCTCTCCAGGACACTGATGSCACAGCGGGAAGAACAGTGTATGA 911
234 eTYrTYrSerProGluArgGlyLeuGlnPhelLeuThrAsnThrCYsTP 250
|||||
912 GCACCTACTCCCTGAGAGAGGCTCCCAATTCCTGCTAATACGCTCTGG 961
251 AsnGlyCYsSerThrProAspAlaGluSerSerLeuAspGluAlaValAr 267
|||||
962 AATGGGTGCTCACCTCTGATGAGAGACTCTCTGATGAGAGACTCTGAG 1011
267 gValProSerSerSerAlaSerGluAsnAlaIleGlyAspProTYrLYst 284
|||||
1012 AGTTCCATTAAGCTCCGCTCGAGAGAACTATTGTTGATCCATATATAA 1061
284 hNisIleSerSerAlaAlaLeuSerLeuAsnSerLeuHISThrValArg 300
|||||
```

```
1062 CTCACATTTCCAGTGCAGCGCTGTCTCTGATTCCTTGCACTACTGTCAGA 1111
301 TYRTYrSerLYsValThrProSerTYrAspAlaThrAlaValThrAsnG 317
|||||
1112 TATTATTCTAAAGTACACCCCTTCTACATGCACTGCTGTAACAAATCA 1161
317 nasIleSerIleHISLeuProSerAlaMetProProValSer_SerPhe 333
|||||
1162 GAACATTTCAATTCAATTACTTTCAGCCATGCCCCCGGTCTGAAAGCTTC 1211
334 SerLeuGlyTYrAlaLeuLeuGlyTYrAlaArgArgAlaLeuHISlePr 350
|||||
1212 TCCCTCGGACTGCACTTCTGGTTGGGCCAGACGTCACATACATGCC 1261
350 oThrValCYsAsnSerPheGlyArgIleLYsSerThrCYsLeuLYsPheT 367
|||||
1262 AACTGCTCTCAACAGTTTGGCGCTATTAAAAGCACATGTTGAAATTCA 1311
367 hTLeuSerThrTYrTYrTPAlaGlnPhAspAsnLeuGlyLYsValGlu 383
|||||
1312 CACTCTCAACCACTTACTGGCGCAGTTTGACAAATCTGGAAGAGTGGAA 1361
384 GlnArgMetIleLeuLYsAlaProProLYsAspLeuIleSerLYsgLude 400
|||||
1362 CAAAGATGATTTTGAAGCTCCACCCAAAGCCTAATATCAAAAGAGTT 1411
400 uAlaTPheGlyPhe 405
|||||
1412 GGCATGTTTGGCTTC 1427
```

seq_name: gb_hlg20:AL160392

seq_documentation_block:
LOCUS AL160392 158888 bp DNA HMG 16-SEP-2000
DEFINITION Homo sapiens chromosome 13 clone RP11-121N13, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL160392
VERSION AL160392.10 GI:10185503
KEYWORDS HMG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 158888)
Tromans A.
Direct Submission
Submitted (15-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 18, 2000 this sequence version replaced gi:10039688.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA121N13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 156753 bases at least Q40
Consensus quality: 156783 bases at least Q30
Consensus quality: 156783 bases at least Q20
Insert size: 158788; sum-of-contigs
Insert size: 146321; 14.5% error; agarose-fp
Quality coverage: 8.48x in Q20 bases; sum-of-contigs Quality
coverage: 9.34x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 156788: contig of 156788 bp in length
* 156789 156888: gap of 100 bp
* 156889 158888: contig of 2000 bp in length.
Location/Qualifiers
1. 158888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-121N13"
/clone_lib="RPC1-11.1"
1. 156788
/note="assembly_fragment:03393"

FEATURES

source

misc_feature
clone_end:SP6
vector_side:right
clone_end:T7
vector_side:left
156889..158888
/note="assembly_fragment:03390"misc_feature
BASE COUNT 51320 a 30978 c 30194 g 46296 t 100 others

ORIGIN

alignment_scores:

Quality:	443.50	Length:	666
Ratio:	2.899	Gaps:	4
Percent Similarity:	22.973	Percent Identity:	21.321

alignment_block:

US-09-389-000-2 x AL160392/rev ..

Align seg 1/1 to reverse of: AL160392 from: 1 to: 158888

243 GlnPheLeuThrsAnthrCysTrpAsnGlyCysSerThrProAspAla1 239

56866 CACTTATGTCGAAT.....AACAGTCG 56844

259 userSerLeuAspGluAla.....V 266

56843 TTCTGAATTACAGAAATGTGATAGTACCCAAATTTGTTCTTCTCA 56794

266 alArgValProSerSerSerAlaSerGluAsnAlaIleGlyAspProTyr 282

56793 TTAAAGTTCATCAAGCTCCGCTCAGAGAAATGCTATTGGTGATCCATAT 56744

283 LysThrHisIleSerSerAlaAlaLeuSerLeuAsnSerLeuHisThrVa 299

56743 AAAATCACAATTCAGATGCAGCGCTGCTGTAATCCTTGCAATCTGT 56694

299 larGlyTyrSerLysValThrProSerTyrAspAlaThrAlaValThra 316

56693 CAGATATTATTCTAAAGTCAACCCCTTCTAGAGCAACTGCTGTAAACA 56644

316 snGlnAsnIleSerIleHisLeuProSerAlaMetProProValSer.Se 332

56643 ATCAGAAATTCATTCAATTCATTACCTTCAGCCATGCCCGGCTCTCAAG 56594

332 rPheSerLeuGlyThrAlaLeuLeuGlyTyrPalaArgAlaLeuHis1 349

56593 CTTCTCCCTCGCAGCATGCTCTGGGTGGCCAGAGCTGACTACACA 56544

349 leProThrVal..... 352

56543 TCCCACTGTCTGTAGATTTCAGGGGTCCATGATGATGAGACCT 56494

352 352

56493 TTCACGCCAGCAGAAACCCCTGACCTGCATTTGTTGCTACTCT 56444

352 352

56443 GGTTGAGAGCTGGACCTAAGAGCTTGATTCTGAGTCTCTACCTGA 56394

352 352

56393 TGAGTAAAGAGGTGCAGTCAACCCATGCTCTGCTCTATGCTGA 56344

352 352

56343 GGGCTGTGTCTGATATCCAGAACCAAGTGAAGCTGATGGAGCTGTCT 56294

352 352

56293 AGTTGACGGTACAGAGTGAAATCTGATCTCCCTAGCAGTATTGCC 56244

352 352

56243 ACCCTCTGCTTCTACTCAGAGCTTCCACAGTGGCACATCCATTTAGC 56194

352 352

56193 ACCCTATCTCTCACTGGCCCGTACTACTGCTACTTATCTTAGGT 56144

352 352

56143 GGGCAGCATCCTTACACTAATCAAAAGCTTCTTCAGTACACAGCA 56094

352 352

56093 GGCTATTTCAGATACAGAGTGTCTTCATACAGAGATATTACTGACA 56044

352 352

56043 CCCTATATGTTCTGACATTGAGTTTGTCATTGTACTAAGATGTGTATA 55994

352 352

55993 CAAAGTAGACAGATCCCTGCCCTCATGACAGATTTAGTTAAGGGAT 55944

352 352

55943 GCAGACAAAGACAGAAATTTACATATGTTGAACGGGTAGAGAGT 55894

352 352

55893 GCCCACTATATATATCAGATAGTGGCAAGTTAGTACTGCTAC 55844

352 352

55843 CCCTGAGAAATACATCTATGTCAATCTCTAGCCACTAAGAAATAGCTA 55794

352 352

55793 GTTAGAGGGGTGAGATGGAGTAGTAGCAAGACAGCTCTGCGAG 55744

352 352

55743 ATGGGTACAGATATGAACAGTACATACCTGATTTGTTGGAAGACTGTC 55694

352 352

55693 TAGGTAGAGAGAGACAACTAATTTGGCTTAAATGAGACTTC 55644

352 352

55643 TGTTAATTTAATATAGTTGCTAATCTTTGCTATTTAACAATAGCTAA 55594

352 352

55593 CATGTTTGGACACTCATATACTAATAGTCCAAATAGTCTGTACTTA 55544

352 352


```

misc.feature      155874..187508
                  /note="assembly_fragment:01108
                  fragment_chain:2
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      40496 a 54492 c 52718 g 38998 t 804 others
ORIGIN

```

```

alignment_scores:
  Quality: 129.00      Length: 413
  Ratio: 0.701        Gaps: 22
  Percent Similarity: 44.552  Percent Identity: 23.002

```

```

alignment_block:
US-09-389-000-2 x AL158822/rev ..

```

```

Align seg 1/1 to reverse of: AL158822 from: 1 to: 187508

```

```

31 SerLeuPheValPheIleIleProGluAsnPhelysGlyCysIleSerG1 47
   ::::: ||||| :::::
16454 ACCCTGCTGCTCTTCCACACCGCAA.....AGCAG 16423
   :::::
47 yHISGlyMetAspIleAlaLeuThrGluProLeuThrMetGluMetS 64
   :::::
16422 CCATGGTTGGGCTGGC..... 16404
   :::::
64 eAsnValValLysTyrTrpThrCysProSerAsnThrValLysThr 80
   |||||:::
16403 .....TGGACAGCTGTCCGAGGTCACAGGGGTCAGG 16371
   :::::
81 GluAsnAlaThrGlyProGluGluLeuProLeuGlnArgSerTy 97
   :::::
16370 CCCAGGCTGTGCTGATGATCGAGTCGAGGCT..... 16338
   :::::
97 rSerGluHISLeuGlyTyrPheProThrAspLeuPheAlaCysSerGlu 113
   |||||:::
16337 ..TCAGACCAAGGGGGG.....AGCTTGATCATCTTGTAGCCCC 16298
   :::::
114 ..SerLeuArgAsnGlyAsnGlyLeu.....GluLeuAsn 124
   |||||
16297 TCTGCTGACACAGCCCAAGGCTCTATCTGGCAGTTCTAGCAGTTGCTG 16248
   :::::
125 AlaSerLeuSerGluPheGluLysAsnLysLysIleSerLeuLeuHis 141
   :::::
16247 TCCCTCTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 16198
   :::::
141 eSerLysGluLysLeuArgArgGluArgGlyLeuLysCysGluGln 157
   :::::
16197 GAAC.....AGGAGCGGATGTCGTTGACCTGTGAGCGT 16163
   :::::
158 LeuArgThrLeuLeuProTyrValLysArgLysAsnAspAlaAla 174
   :::::
16162 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 16113
   :::::
174 rValLeuGluAlaThrValAspTyrValLysTyrIle..... 186
   |||||
16112 GGTCTGAGAGATGTCGTGAGTCTGCGGCTTGCACAGCGCCCTGGGGC 16063
   :::::
187 .....ArgGluLys 189
   :::::
16062 CCAGTCAGAGCAGCAGCTGTGAGTTGAAGGTAGACCCCTCAGTCC 16013
   :::::
190 IleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
   :::::
16012 TCCGGATACGCTGATGGACATCTCACC..ACCTTGAGGCA.... 15971
   :::::
206 tArgPheCysLysLysGlnInThrProIleGluLeuSerLeuProGlyT 223
   |||||
15970 .....GTGTG.....CCCCGTCATCCCCCTGCTCCTTACAG 15940
   :::::
223 hr.....ValMetAlaGlnArgLysAsnSerValMetSerThrTyrSer 237
   :::::

```

```

15939 CCTCTGGGCTCCACGGCTAGACGAGCCAGCTTTTGGCAGGGCAGT 15890
238 ProGluArgGlyLeuGlnPheLeuThrAsnThrCys...TrpAsnGlyC 253
   |||||
15889 CCGATTGTTCCAGGCTTACTGTTGATTCCTGCTCCGTAAGAGAC 15840
   :::::
253 sSerThrProAspAlaGluSerSerLeuAspGluAlaValAlaVal... 268
   :::::
15839 AAGCCTTCTCTCCAGCCCTCTGCGCCCTGATGCTTCTGATGAGG 15790
   :::::
269 .....ProSerSerAlaSerGluAsnAlaIleGlyAsp 280
   |||||
15789 GTTCCGAGATCTCCGTCCCAATTGGGGAGATGGGCTCTCTGAT 15740
   :::::
281 Pro.....TyrLysThrHisIleSerSerAl 289
   |||||
15739 CCTCCCTGGCTGAGCACCCTAGATCATTCCTCTGACGCTCCACAG 15690
   :::::
289 AlaLeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysVal 306
   :::::
15689 AAGAGCTCTCTCCCTCCCAAGGCTGCGGGCTCTGG.....CTGA 15646
   :::::
306 hrProSerTyrAspAla.....ThrAlaValThrAsnGlnAsnIleSer 320
   |||||
15645 CCCCCTGAGTCTGGGGCTGAGGACCTGTGTGGGAAATGAATCTGTG 15596
   :::::
321 IleHisLeuProSerAlaMetProValSerSerPheSerLeuGlyTh 337
   :::::
15595 TCATCTGTCAGGTCAGGAGGCTCCAGCATTTGCAGTGAGGCCCTTG 15546
   :::::
337 rAlaLeuLeu.....GlyTrpAlaArgArgAlaLeuHisIlePro 350
   |||||
15545 GCACAGCCCTGATTCAGAAATGTTGGGCTTTCGGGGTGAACCTCAG 15496
   :::::
351 ThrValCysAsnSerPheGlyArg..... 358
   |||||
15495 CTCCTTGTACCTGATCTCTGCTGCCCCCAAGTCCAGAGGGGAGCCG 15446
   :::::
359 .....IleLysSerThrCysLeu 365
   |||||
15445 TCACAGTACCTCCAGGCCCTGGGGTAGCAGTCGTACAGGCTGTCTG 15396
   :::::
365 yAsPhe.....ThrLeuSerThrThrTyr 372
   |||||
15395 GGTTCCTCAGCTGATGTCGTCTGCTGTTGTTTAC 15361
   :::::
seq_name: gb_hhg20:AL354761
seq_documentation_block:
LOCUS      AL354761      172727 bp      DNA      06-SEP-2000
DEFINITION Homo sapiens chromosome 9 clone RP11-98L5, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION      AL354761
VERSION      AL354761.6 GI:10039788
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 172727)
AUTHORS      Plumb,B.
TITLE      Direct Submission
JOURNAL
Submitted (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9800828.
COMMENT
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: Ba98L5

```

```

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 169466 bases at least Q40
Consensus quality: 170480 bases at least Q30
Consensus quality: 171223 bases at least Q20
Insert size: 171837; sum-of-contigs
Insert size: 180951; 1.3% error; agarose-IP
Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
Coverage: 8.40x in Q20 bases; agarose-IP
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
9670 9769: contig of 9669 bp in length
9770 24788: contig of 15019 bp in length
24789 24888: gap of 100 bp
24889 45348: contig of 20460 bp in length
45349 45448: gap of 100 bp
45449 113889: contig of 68441 bp in length
113890 113989: gap of 100 bp
113990 127156: contig of 13167 bp in length
127157 127256: gap of 100 bp
127257 131636: contig of 4380 bp in length
131637 131736: gap of 100 bp
131737 146937: contig of 15201 bp in length
146938 147037: gap of 100 bp
147038 153346: contig of 6309 bp in length
153347 153446: gap of 100 bp
153447 159184: contig of 5738 bp in length
159185 159285: gap of 100 bp
159285 172727: contig of 13443 bp in length.
Location/Qualifiers
1..172727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-98L5"
/clone.lib="RC11-11.1"
1..9669
/note="assembly_fragment:02116
fragment_chain:1"
9770..24788
/note="assembly_fragment:01142
fragment_chain:1"
24889..45348
/note="assembly_fragment:02183
fragment_chain:1"
45449..113889
/note="assembly_fragment:02357
fragment_chain:2"
113890..127156
/note="assembly_fragment:01763
fragment_chain:2"
127257..131636
/note="assembly_fragment:03395
fragment_chain:2"
131737..146937
/note="assembly_fragment:02045
fragment_chain:3"
147038..153346
/note="assembly_fragment:01090
fragment_chain:3"
153447..159184
/note="assembly_fragment:00178
fragment_chain:3"
FEATURES
source

```

```

misc_feature 159285..172727
/note="assembly_fragment:01839
fragment_chain:3
clone_end:17
vector_side:right"
BASE COUNT 40126 a 45744 c 46532 g 39420 t 905 others
ORIGIN

alignment_scores:
Quality: 123.50 Length: 417
Ratio: 0.660 Gaps: 24
Percent Similarity: 44.844 Percent Identity: 23.261

alignment_block:
US-09-389-000-2 x AL354761/rev ..

Align seg 1/1 to reverse of: AL354761 from: 1 to: 172727

31 SerLeuPheValPheIleIleProGluAsnPhelGlycylSerG1 47
||||| ||||| |||||
98372 ACCCTCGCTGCTCTTTCCACACCGCA.....AGCAG 98341
47 YHISGlyMetAspPILeAlaLeuThrGluProLeuThrMetGluLysMets 64
||||| ||||| |||||
98340 CCATGCTTTGGCGCTGGC..... 98322
64 eAsnValValLysTYTTPThrThrCysProSerAsnThrValLysThr 80
||||| ||||| |||||
98321 .....TGACAGCGCTGTGGCAGGTCCAGGCGTCAGG 98289
81 GluAsnAlaThrGlyProGluLysLeuGlyLeuProLeuLysArgSerTy 97
|||
98288 CCAGGCTGTCTGATGTGATCTCGGCGCT..... 98256
97 rSerGluLysLeuGlyTYrPheProThrAspLeuPheAlaCysSerGlu 113
||||| ||||| |||||
98255 .TGAACACAGGGGGC.....AGCTTGAGTCATCTTGAGCCCC 98216
114 ..SerLeuThrAsnGlyAsnGlyLeu.....GluLeuAsn 124
|||
98215 TCCTCGCTGACAGCGCCAGGCGTGTATCTGCGAGCTTACAGAGTCTG 98166
125 AlaSerLeuSerGluPheGluLysAsnLysLys..IleSerLeuLys 141
||||| ||||| |||||
98165 TCCCTTCTGGGGAGGGGCAAGAGGCCCAAGATATTCTTCTTGACCG 98116
141 eSerLysGluLysLeuArgGluArgLysLysTyrcysCysGluGln 157
||||| ||||| |||||
98115 GAAC.....AGAAAGCGATGCTGAGCTGAGCGT 98081
158 LeuArgThrLeuLeuProTyValLysGlyArgLysAsnAspAlaLase 174
||||| ||||| |||||
98080 CTGGCGGCGCTGCTGCCAGTTCGATGCGCGGAGAGACATGAGCTTC 98031
174 rValLeuGluAlaThrValAspTyValLysTyrl..... 186
||||| ||||| |||||
98030 GGTCTGAGATGTGTGACATCTCGGCTGCCAGCCCGCTGGGGC 97981
187 .....ArgGluLys 189
97980 CCAGTCAGAGACAGCAGCTGATGTGAAGGTAGACCGCTGATGTC 97931
190 IleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMe 206
|||
97930 TCCGGGCGTACGTCATGGACATCTCAC..ACCTTGACAGCA..... 97889
206 tArgPheCysLysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
|||
97888 ....GTGCT.....CCCTGTCATCCCGCTCTCTGAG 97858
223 hr.....ValMetAlaGlnArgGluAsnSerValMetSerThrTySer 237
|||

```


629 CTGCGCGCGCTGGCGCCGCCCTCTGCTAGGCCCGCTGCT 678
 293LeuAsnSerLeuHisThrValArgTyrTyrSerLysVal 305
 679 CCTGGCGCTGCTGCCCGCCAGCGCGTGTCAAGCCCTCAGT..... 722
 306 ThrProSerTyrArgPalaThrAlaValThrAsnGlnAsnIle..... 319
 723 TCTCCTTCAAGTTAGCGCAGCGCTCTCCAGCGCCCACTGCTGTCATTC 772
 320SerIleHisLeuProSerAla.... 326
 773 TGGAGCTCCAGCTGCTGCTGGCGCTGCTGCGACCTGCCCTTGGCGGTC 822
 327MetProValSerSerPheSerLeuGlyThrAlaLeuLeuGly 341
 823 AGGCGCTGCGAGCTGCTGCCCTCTCCAGCTGGGACAGTCCCTCGCAA 872
 342 TrpAlaArgAlaLeu...His.....IlePr 350
 873 GAGGCGAGGCGCCAGCTTCCACATCCGAGCCCTGTGACATAGCGGCC 922
 350 ThrValLys.....AsnSerPheGlyArgGlyLeuSerThrC 363
 923 CACGGTCTGTTCTCAGATTCTATCATTCACAGAGATTAAAGTCATTT 972
 363 yslLeuLys 365
 973 GCTGCCAA 980

seq_name: gb_pr3:AF070992

seq_documentation_block:

LOCUS AF070992 2313 bp mRNA PRI 10-SEP-1999
 DEFINITION Homo sapiens HPV-16 E2 binding protein (E2BP-1) mRNA, complete cds.
 ACCESSION AF070992
 VERSION AF070992.1 GI:5852278
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 1 (bases 1 to 2313)
 AUTHORS Zheng, P.-S. and Pater, A.
 TITLE Submitted (08-JUN-1998) Division of Basic Science, Faculty of
 Medicine, Memorial University of Newfoundland, St. John's,
 Newfoundland A1B 3X6, Canada
 JOURNAL Location/Qualifiers

FEATURES
 source 1..2313
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

gene 1..2313
 /gene="E2BP-1"

CDS 76..1434
 /gene="E2BP-1"

protein_start=1
 /product="HPV-16 E2 binding protein"

protein_id="AAD53986.1"

translation="MTEVEYTOLOHILCSHNEAANDELRLNSALLAAGPAGAG
 GFAAGCGGAGAPVPLCPALADAPLGHIDLELMMLSPAGAAETSGGDA
 RARADGAKGAGAGAAAGADGAPARAVRREDRNSIPAPPPAPRGPDP
 EPGALNNITLTPPELMNVPCOONKCALVKNKTAATTTLQETPLPTNAS
 TSGNSNLSQTOSSNSCVLEAKHDIQIPRAFSTYQOETESTKTLDSKRVLE
 QVNLKVEALCKQALKRNRMRQDITNERRALGEIQNGEATVQAGWVSES
 QANLGEQAGSGPQGRSQRRENNRMRERRRIRICDELNLVPCNAETDATT
 QMTTAFLLKTYIOERHGDLSLKEFESEVFCGKTRRLKLRPSLVTCPAGSLQSSPME
 IK"

misc_feature 124
 /gene="E2BP-1"

note="alternate position for initiation methionine"

BASE COUNT 608 a 558 c 609 g 538 t

ORIGIN

alignment_scores:

Quality: 119.00 Length: 101
 Ratio: 1.776 Gaps: 3
 Percent Similarity: 66.337 Percent Identity: 28.713

alignment_block:

US-09-389-000-2 x AF070992 ..

Align seg 1/1 to: AF070992 from: 1 to: 2313

140 HisSerSerLysGluLysLeuArgArgGluArgIleLysTyrCysGly 156
 1144 CATACCGAATGGAAGAGATGAGAGCCGACAGATCCGATTGCTGTA 1193
 156 uGlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAla 173
 1194 TGAGTGAATCTCTTATGTCGCTTGTGCAATGCC...GAGACTGACAAAG 1240
 173 lAserValLeuGluAlaThrValAspTyrValLysTyrIleArgGluLys 189
 1241 CCACACTCTGCGAGTGGACACGATTCCTGAAATACATCCAGGAAAG 1290
 190 lIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
 1291 CATGCGATTCCTTAAAGAAATTGAGACGTA..... 1326
 206 lArgPheCysLysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
 1327 ...TTTTCGGTAAAGCTGCGGAGGCTAAAGCTGACCAAGCGGACT 1372
 223 hr..ValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerPro 238
 1373 CTTGGTGAAGCTCTCTGCGAGGAGGAGTTTACAGAGAGAGCCCTCGATG 1422
 239 Glu 239
 1423 GAG 1425

seq_name: gb_pr1:AB012124

seq_documentation_block:

LOCUS AB012124 2316 bp mRNA PRI 09-JAN-1999
 DEFINITION Homo sapiens TCFL5 mRNA for transcription factor-like 5, complete
 cds.
 ACCESSION AB012124 GI:4126408
 VERSION AB012124.1
 KEYWORDS transcription factor-like 5; TCFL5.
 SOURCE Homo sapiens cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Homidae; Homo.

REFERENCE
 1 (bases 1 to 2316)
 AUTHORS Nakamura, Y., Nishimori, H. and Maruyama, O.
 TITLE Direct Submission

JOURNAL Submitted (12-MAR-1998) to the DDBJ/EMBL/Genbank databases, Yusuke
 Nakamura, Institute of Medical Science, The University of Tokyo,
 Laboratory of Molecular Medicine, Human Genome Center; 4-6-1
 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
 (E-mail:yusuke@ims.u-tokyo.ac.jp, Tel:81-3-5449-5372,
 Fax:81-3-5449-5433)

REFERENCE
 2 (sites)
 AUTHORS Maruyama, O., Katagiri, T., Nishimori, H., Miki, Y., Ueno, A. and
 Nakamura, Y.

TITLE Cloning, mapping, and genomic organization of THL1, a novel human
 testis-specific gene containing a basic Helix-Loop-Helix motif

JOURNAL Unpublished (1998)
 COMMENT Sequence updated (27-Jun-1998).
 FEATURES Location/Qualifiers

1..2316
 /organism="Homo sapiens"


```

137 erLeuLeuHisSerSerLysGluLysLeuAArgGluArgIleLysTyr 153
      ::::::::::::::::::::|||
504 .....AACGCCGAGAGCGCACCGCGTGAACAGTGAACAAC 542
154 CysCysGluGlnLeuArgThrLeuLeuPro.....Tyr 164
      ::::::::::::::::::::|||
543 AGCTTCCGCGCTGTGGCCAGACATTCGCCATGATCATCGCTGACCT 592
164 rValLysGly.....ArgLysAsnAspAlaIaIaSerV 175
      ::::::::::::::::::::|||
593 GACCAAGAGGAGTGGCGGGGCGGCACAGAAAGATCTCCAAGTGGACA 642
175 alLeuGluAlaThrValAspTyrValLysTyrIleArgGluLysIle... 190
      ::::::::::::::::::::|||
643 CACTGGCGATTGCCGTGAATACATCCGCGCGCTGACAGACCTGCTGCAC 692
191 .....SerProAlaValMetAlaGlnIleThrGluAlaIle 202
      ::::::::::::::::::::|||
693 GACCTCAATGGCGGTGCCCGCCAGCTCCAGTGGCGCGCGAGTGGCT 742
202 uGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleGluLeu 219
      ::::::::::::::::::::|||
743 CGTCAGCGACGCTGCAGCTCTGCTGGAGAGAGAAC..... 779
219 erLeuProGlyThrValMetAlaGlnArgLysAsnSerValMetSerThr 235
      ::::::::::::::::::::|||
780 .....TGCAGCAACAGCTCCAGCAGCTCC 803
236 TyrSerProGluArgGlyLeuGlnPheLeu..... 245
      ::::::::::::::::::::|||
804 GCGAGCGACGACCATGCGACGACCTCTACTACAGCAGCTCCCTGCC 853
246 .....ThrAsnThrCysTrpAsnGlyCysSerThrProA 257
      ::::::::::::::::::::|||
854 CGCAGCGCGGAGCAGCAGCAGCAGATCAGCGCGGAGCTACACAG 903
257 spAlaGlnSerSerLeuAspGluAlaValArgValProSerSerSerAla 273
      ::::::::::::::::::::|||
904 ACGAGAGCTCC.....AACTCCAACTCC 926
274 SerGlnAsnAlaIleGlyAsp.....Pr 281
      ::::::::::::::::::::|||
927 AACTCCAACTCCCTGCAGACAGCAGCAGCTACAGCCCGCAGATTCC 976
281 oTyrLysThrHisLeuSerSerAlaLeuSerLeuAsnSerLeu... 296
      ::::::::::::::::::::|||
977 CCATCAGCAACACTGACTGCTGCTGTAACAGCAACTGATCGGCG 1026
297 .....HisThrValArg 300
1027 TGGGTGCAGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1076
301 TyrTyrSerLysValThr.ProSerTyrAspAlaThrAlaValThrAsn 316
      ::::::::::::::::::::|||
1077 TGGCACTTGGCCACTCTCTCTTCACTCGAGCATGTCTTCTGACTCGGG 1126
317 .....GlnAsnIleSerIle...HisLeu..... 323
      ::::::::::::::::::::|||
1127 CACCTATGAGCGCGCGCCGACGAGCTGTGCGACCCGACATGAGGTG 1176
324 .....ProSerAlaMetProPro...ValSerSerPheSe 334
      ::::::::::::::::::::|||
1177 GACATGGAAGTGAAGCGGAGCGGATGCCGCCAGCTGACGCTCAAGTTC 1226
334 rLeuGlyThrAlaLeuLeuGlyTyrPAlaArgAlaAlaLeuHisIleProT 351
      ::::::::::::::::::::|||
1227 GAGCCCTACGAGCACTTCCAGCTGAGAGAGAGAGACTGACAC...CCGA 1273
351 hrValCysAsnSer..... 355
      ::::::::::::::::::::|||
1274 CGACGAGAGAGATCTCGACTACATCTCCATGTGGCAGAGCATGATCAT 1323
356 .....PheGlyArgIleLysSerThrCysLeuLysPheThrLeuSe 369

```

```

seq_name: gb_ov: XELMAX4A
seq_documentation_block:
LOCUS      XELMAX4A      878 bp      mRNA      VRT      06-APR-1994
DEFINITION Xenopus laevis XMax4 mRNA, complete cds.
ACCESSION  L09739
VERSION    L09739.1 GI:214912
KEYWORDS   Myc protein; variant.
SOURCE     Xenopus laevis (library: Lambda gtl1 (by C. Kintner) embryo
           neuration stage cdna to mRNA.
ORGANISM   Xenopus laevis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
           Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 878)
AUTHORS   Tonissen, K.F. and Krieg, P.A.
TITLE     Analysis of a variant Max sequence expressed in Xenopus laevis
JOURNAL   Oncogene 9 (1), 33-38 (1994)
MEDLINE   94134435
FEATURES
  source
    1..878
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    1..110
    111..602
    /standard_name="XMax4"
    /function="forms heterodimers with the Myc protein family"
    /protein_id="AA17425.1"
    /db_xref="GI:214913"
    /translation="MSDNDIDVESDESDSRFPYSADKRAHNNALERRRDKIKSFH
    GURDSVPSLOGSKASRAQIILDKATEYIOYMRKRNHQQDIDILKRNALLDQVOIS
    NRPSPNGAPPEOKNVLQILSPKVRALERAKSSSLOQSNYSSESETEBPQSRKRLM
    DAS"
  3'UTR
    BASE COUNT      252 a      228 c      163 g      235 t
    ORIGIN
alignment_scores:
  Quality: 116.50      Length: 302
  Ratio: 0.793      Gaps: 13
  Percent Similarity: 48.675      Percent Identity: 21.854
alignment_block:
US-09-389-000-2 x XELMAX4A
Align seg 1/1 to: XELMAX4A from: 1 to: 878
86 ProGluGluLeuGlnLeuProLeuGlnArgSerTyrSerGlnHisLeuG1 102
      ::::::::::::::::::::|||
66 CCGGAGCAGCAGCCTCTCC..... 86
102 yTyrPheProThrAspLeuPheAlaCysSerGluSerLeuArgAsnGlyA 119
      ::::::::::::::::::::|||
87 .TCCTTCCA.....TCACTATAGAAAGATGAGCATATACG 123
119 snGlyLeuGluLeuAsnAla.....SerLeuSerGluPheGluLysAsn 133
      ::::::::::::::::::::|||
124 ATGACATCGAGGTGAGAGCAGCAAGACTGCTCAAGATTCCTTATTCG 173
134 LysLysIleSerLeuLeuHisSerSerLysGluLysLeuAArgArgGluAr 150
      ::::::::::::::::::::|||
174 GCAGACAAACAGCGCCATTCACAAATGCACTGAGCGGAACGACGAGACA 223
150 gIleLysTyrCysGluGlnLeuAArgThrLeuLeuProTyrValLysG 167

```

[illegible]

seq_name	seq_documentation_block	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE
gb_cco:MUSHSB86a		MUSHSB86a	2975 bp mRNA					
		Mouse heat shock protein 86 mRNA, complete cds, and 28S ribosomal RNA, partial sequence.	10-NOV-1994					
		J04633						
		J04633.1	GI:194030					
		heat shock protein.						
		Mouse (Strain BALB/c)	CDNA to mRNA.					
		Mus musculus						
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
		Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus						
		(Pases 1 to 2975)						

AUTHORS	MOORE, S. K., KOZAK, C., ROBINSON, E. A., ULLRICH, S. J. and APPELLE, E.
TITLE	Murine 86- and 84 kDa heat shock proteins, cDNA sequences, chromosome assignments, and evolutionary origins
JOURNAL	J. Biol. Chem. 264 (10), 5343-5351 (1989)
MEDLINE	89174568
REFERENCE	2 (bases 1 to 2975)
AUTHORS	MOORE, S. K.
TITLE	Direct Submission
JOURNAL	Submitted (11 APR 1989) Stephen K. Moore, Laboratory of Cell Biology, National Cancer Institute, NIH, Bethesda, MD 20892, USA
FEATURES	Location/Qualifiers
SOURCE	1. 2975

MRNA	<1. .1588
	/note="heat shock protein 86 mRNA"
rRNA	<1. .>229
	/product="28S ribosomal RNA"
CDS	261. .2462

BASE COUNT 959 a 549 c 719 g 748 t
ORIGIN

```
alignment_scores:      Length:    3811  
                        Quality:    115.50  
                        Ratio:       0.628  
Percent Similarity:    48.294     Percent Identity: 22.835
```

```
alignment_block:
US-09-389-000-2 x MUSHSP86A .
```

Align seg 1/1 to: MUSHSP86A from: 1 to: 2975

```

6 ValPSeSerIeuSerzIaIuIuIeuGuaIaIaIeIyIeuIleAr 22
:::||||| :::::||||| ::||| ::|||
1443 CTCTCTCTAAATATTTCCCGTAAATGCTGCACAAAGTAATTAATTCGAA 1492
22 gPhegIyLys.....LysAsnThrIseSerLeuPheValPheI 36
:::||||| ::||| ::||| ::||| ::||| ::|||
1493 AGTTATCAAGAAAGAAATTTGGTCAGAAAGATGCTTAGCAATTTACTGAC 1542
36 IeLe .....ProGuaIaIuPheIySgIyCysIleSeCdyIhISgIyMet 50
::: ||||| ::||| ::||| ::||| ::|||
1543 TACACAGACATAAAGAGAACTACAAAAGATTATAGACCACTGTCAAAA 1592
51 AsPIaIaIeIu.....ThrCluProIeuThrMetgIyIyMetSe 64
:::||||| ::||| ::||| ::||| ::|||
1593 AATATTAAGCTTGGAAATTCACAGAGACCTCTAGAAATCGAAGAAAGCTTTC 1642
64 IaSnValValLySTyTTPThThCysProSerAsnThrValIySThrg 81
:::||||| ::||| ::||| ::||| ::|||
1643 AGAGCGTGTGGCGTACTGACAC..... 1664

```

81	lulsalrhrgrProgluIngluEgluLeu.....Probu	93
1665	..TCTGCTTCTGGGAGACAGATGGTTTCTCTGAAGACTACTGTACAGA	1712
94	glnArSerTyrSerGluHisLeuGluTyrPheProHr.....	106
1713	ATGAAAGAAACCAAGAAAGCATC...TATTATTACAGAGTGAACCA	1759
107	AspLeuPheIlaCysSer.....GluSerLeuArgasnGluAsn	120
1760	GGACAGGTGGCTAACTCCGCTTGTGGAAAGCTCCGAAG...CATG	1806
120	lyuEduLeu.....AsnAla	125
1807	GCATTAACTAATTATTATGATTCAGACCCATGATGACGATTTGTGCA	1856
126	SerLeuSerGluPheGluLysasnLysLleSer.....	137
1857	CACCTGAAGAAATTTTGGGCAAGACCTTGTGTGTATTACCAAGAAG	1906
138IeuLeuHisSerSerLysLysLeuArgHrGlnArgIle	152
1907	ACTGGAACCTCCAGAGATGAAGAGAAAGAAAGAAACAGAGAGAA	1956
152	ysTyrCysCysGluGlnLeuArgThrLeuLeuProTyrValLysGluArg	168
1957	AGACAAATTTTGGACCTGTGCAAAATATG.....	1988
169	LysAsnAspAlaIlaSerValLeuGluAlaThrValAspTyrValLysTy	185
1989AAATATTTTGGAGAAAGGTGAAAGCTGGTTGT	2026
185	rIlaeArgLysIle...SerProAlaValMetAlaGlnIleThrGlu	201
2027	GTCAAAACCACTGGTACATCCCCGGCTGTATTGTCAACACATATG	2076
201	IaLeuGlnSerAsnMetArgPheCysLysLysGlnIleThrProIleGlu	217
2077	GGTGGACAGCAAAACATGGAGATCATGAAAGCTCAA.....	2144
218	LeuSerLeuProGluThrValMetAlaGlnArgLysAsnSerValMetSe	234
2115GCCCTCAAGACAACTCAACATGGG	2140
234	rThrTyrSerProGluArgLysGlnPheLeuThrAsnThrCysTrpA	251
2141	TTTACATGGCAAAACAACTGTGAGTA.....	2171
251	snGlyCysSerThrProAspAlaGluSerSerLeuAspGlnAlaValArg	267
2172AATCTGTAT.....CACTCATATTGAAACCTTAAG	2204
268	ValProSer...SerSerIaSerGluAsnAlaIleGluAsp.....	280
2205	CAAAAGCGACAGGCTACAAAGATGCAAAATGTGTGAAGATCTGTAT	2254
281	...ProTyrLysThrHisIleSerSerAlaIaLeuSerLeuAsn...	294
2255	CTGTGCTGTGAACACCTCACTGTCTGTGGCTTCACTGTGAAGATC	2304
295	SerLeuHisThrValArgTyrTyrSerLysVal.....	305
2305	CCGAGACCCATGTTAAACAGATCTACAGATGTAACCTGTGTAGCT	2354
306ThrProSerTyrAspAlaThr.....AlaValThrAs	316
2355	ATTGATGAGGATGATGCTCTATGCTGGATGACACACAGCTGCTGTAAG	2404
316	ngIaAsnIleSerIleHisLeuProSerIlaMetProVal	330
2405	AGAA.....ATGGCTCCCTGG	2420

```

seq_name: gb_pr3:AF040963

seq documentation block:
LOCUS      AF040963          879 bp      mRNA          PRI          20-JAN-1998
DEFINITION Homo sapiens Mad4 homolog (Mad4) mRNA, complete cds.
ACCESSION  AF040963
VERSION    AF040963.1
KEYWORDS   GI:2792361

SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  1 (bases 1 to 879)
AUTHORS   Pribill,I., Barnes,G.T., Chen,J., Church,D., Buckler,A.,
            Ambrose,C.M., Bates,G.P., Lehrach H., Gusella,M.J., Duyeo,M.P.,
            Ambrose,C.M., Macdonald,M.E. and Gusella,J.F.
TITLE      Comparison of Exon Trapping and Sequence-based Methods of Gene
            Finding
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 879)
AUTHORS   Pribill,I., Barnes,G.T., Chen,J., Church,D., Buckler,A.,
            Ambrose,C.M., Bates,G.P., Lehrach H., Gusella,M.J., Duyeo,M.P.,
            Ambrose,C.M., Macdonald,M.E. and Gusella,J.F.
TITLE      Direct Submission
JOURNAL    Submitted (05-JAN-1998) Molecular Neurogenetics Unit, Massachusetts
            General Hospital, 13th street, Charlestown, MA 02129

FEATURES
            source
                1..879
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="4"
                /map="4p16.3"
                1..879
                /gene="Mad4"
                14..643
                /gene="Mad4"
                /note="T72"
                /codon_start=1
                /product="Mad4 homolog"
                /protein_id="AAB97009.1"
                /db_xref="GI:2792362"
                /translation="MEINSLILILEAEVLEERDRBEAGVASYLPEDGPFAREKTKA
            AGLYRKAPNNRSSHNELEKRRRLATRLYLDTOLQVLPIDGSPFHTLISLIRAKVHI
            KKEEDRRALSTKEQLQEQERFLKRLBQLSVQSVRVRTDSTGSAVSTDDSEQVD
            IEGEFEPGEIDSYGSSSDADHDHYSLGTGDSGFPGHCRIGRPALPS"

BASE COUNT      166 a          296 c          285 g          132 t
ORIGIN

alignment_scores:
            Quality: 114.50          Length: 259
            Ratio: 0.867          Gaps: 12
            Percent Similarity: 50.965          Percent Identity: 23.938

alignment block:
US-09-389-000-2 x AF040963 ..

Align seg 1/1 to: AF040963 from: 1 to: 879

131  GULYASnlylslyslleSerleuLeu..... 139
|||||:|||||:|||||:
131  GAGAAACAAAGGCGCGCGCGCTGCGCCAGACGCCCGACACAGACGnc 180
|||||:|||||:|||||:|||||:
140  ...HisSerSetLysGLuLYsLeuArGArGLuArGLyLeIeLYsTYrCYgC 155
|||||:|||||:|||||:|||||:|||||:
181  TTCACACCAACGAGCTGTGAAACACACAGACGCCACCAACGACGCTGTACC 230
|||||:|||||:|||||:|||||:|||||:
155  ySGLuLIuLeuArGTrhLeuLeuPrcTYr...ValLYsGLyArGLySAsn 170
|||||:|||||:|||||:|||||:|||||:
231  TTGAGCAGCTCAAGCAACGCTGTGCCCCCTGGCCCCGACACAGCACCGCCAC 280
|||||:|||||:|||||:|||||:|||||:
171  ASPALALISerValleuGLuAlArhTrValAspTYrValLYsTYrTrLeAr 187
|||||:|||||:|||||:|||||:|||||:

```



```

212 nglnthProIIeGlulLeuSerLeuProGlyThrValMetAlaGlnArg 229
    ::::: ||| |||::: |||
539 CCAAGCTCAGTGAAGGCGCTTGACCCGTGAGACCCACACGCGTGCC 588
    ||| ::::: ||| ::::: |||
229 LuasnerValMetSerThrTyrSerProGlnArgGlyLeuGlnPheLeu 245
    ||| ::::: ||| ::::: |||
589 TTGTCTCTATCTCAGCAGCTGTGCTCCAGCGG..... 623
    ||| ::::: ||| ::::: |||
246 ThrAsnThrCysTrpAsnGlyCysSerThrProAspAlaGluSerLe 262
    ::::: ||| ::::: |||
623 ..... 623
262 uaspgluAlaValArgValProSerSerAlaSerGluAsnAlaIleG 279
    |||::: |||::: |||::: |||
624 ...GAGGACGACGATGACATCCTCCAGCGCCACACCATCACCCCT 669
    |||::: |||::: |||::: |||
279 LysAspProTyrLys.....ThrHisIleSerSerAlaAla 290
    |||::: |||::: |||::: |||
670 TGCACCTCAGCAGCTGGGAGCTGCTTCCACATCTCCCCACAGCCCTG 719
    |||::: |||::: |||::: |||
291 LeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThrPr 307
    |||::: |||::: |||::: |||
720 CTCACGCCCATGACTGCTCCACACA.....TCAGAGTCACACCC 757
    |||::: |||::: |||::: |||
307 oSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuP 324
    |||::: |||::: |||::: |||
758 ATGTCGCTATCCACATCT..... 776
    |||::: |||::: |||::: |||
324 roSerIleMetProValSerSerPheSerLeuGlyThrAlaLeuLeu 340
    |||::: |||::: |||::: |||
777 ..TCAGAGAGCGCTTCTGCTCATGGCTGCTCTCTCCACAGACAGCTTT 824
    |||::: |||::: |||::: |||
341 GlyTrpAlaArgArgAlaLeuHisIleProThr 351
    |||::: |||::: |||::: |||
825 GCCCATGCAAGATCCGCTCTTCGATGCCATCA 857
    |||::: |||::: |||::: |||
seq_name: gb_ro:MMU271867

seq_documentation_block:
LOCUS      MMU271867      2539 bp      mRNA      ROD      11-JUL-2000
DEFINITION Mus musculus mRNA for basic-helix-loop-helix protein (Hey2 gene).
ACCESSION  AJ271867
VERSION    AJ271867.1 GI:6900326
KEYWORDS   basic-helix-loop-helix protein; Hey2 gene.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 2539)
AUTHORS    Leimaster,C., Externbrink,A., Klamt,B. and Gessler,M.
TITLE      Hey genes: a novel subfamily of hairy- and Enhancer of split
            related genes specifically expressed during mouse embryogenesis
            Mech. Dev. 85 (1-2), 173-177 (1999)
JOURNAL    99345954
MEDLINE    2 (bases 1 to 2539)
REFERENCE  2 (bases 1 to 2539)
AUTHORS    Steidl,C., Leimaster,C., Klamt,B., Maier,M., Nanda,I., Dixon,M.,
            Clarke,R., Schmid,M. and Gessler,M.
TITLE      Characterization of the human and mouse HEY1, HEY2, and HEY3 genes:
            cloning, mapping, and mutation screening of a new bHLH gene family
            Genomics 66 (2), 195-203 (2000)
JOURNAL    20318621
MEDLINE    3 (bases 1 to 2539)
REFERENCE  3 (bases 1 to 2539)
AUTHORS    Gessler,M.
TITLE      Direct Submision
JOURNAL    Submitted (31-JAN-2000) Gessler M., Theodor-Bayerl-Institut fuer
            Biowissenschaften (BioCenter), Physiological Chemistry I, Am
            Hubland, 97074 Wuerzburg, GERMANY
FEATURES   1..2539
            location/Qualifiers
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="10"
            /clone_lib="129/ola mouse cosmid"

```

```

gene      /clone="MPCG121M23647Q2"
          136..1155
          /gene="Hey2"
          136..1155
CDS       /gene="Hey2"
          /codon_start=1
          /product="basic-helix-loop-helix protein"
          /protein_id="CAB71346.1"
          /db_xref="GI:6900327"
          /translation="MKRPCEETTSQSDLETDIDVGSNNYPGHATSYMRNSPTTTS
          OIMARKRRGRIEKRDRINISLRLRIVPTAFKGGSAKLEKAEILOWTVHLTK
          LQATGKGYEDAHALADPMSIGRECLTEFARLTSVEGLDPSDPLRVRYSLSLC
          ASQREAVMTSSSMHHHPHPLPHHMAAFHPLPALQPNLHSTSESTPCSLTSSEY
          PSAGSALVATFAHADSLRMPSGGVAPCPVPLSTLSLSTVHAHAATAAAT
          SFPLSFAGAPFMLPSNMAAAVAATAAISPLSVSAASSPOOTSTGNKXPDPWT
          EVGAF"
BASE COUNT 715 a 604 c 553 g 667 t
ORIGIN
alignment_scores:
    Quality: 114.50      Length: 311
    Ratio: 0.774        Gaps: 10
    Percent Similarity: 47.588      Percent Identity: 21.543
alignment_block:
US-09-389-000-2 x MMU271867 ..
Align seg 1/1 to: MMU271867 from: 1 to: 2539
57 ProLeuThrMetGluLysMetSerAsnValValLysTyrTyrThrThrGly 73
    ||||| ::::: |||::: |||
100 CCGCTGCGGACAGTAGCTGCTCTCTCTGTCGCGATGAAGCGCCCTTG 149
    |||::: |||::: |||::: |||
73 sProSerAsnThrValLysThrGluAsnAlaThrGlyProGluGluLeuG 90
    |||::: |||::: |||::: |||
150 TGAGGAACGACCTCCGAAAGCAGCCTG.....GACGAGACCATCG 190
    |||::: |||::: |||::: |||
90 LysLeuProLeuGlnArgSerTyrSerGluHisLeuGlyTyrPheProThr 106
    |||::: |||::: |||::: |||
191 ACGTGGGAGCGAGCAACATTCCTGGGCAC..... 222
    |||::: |||::: |||::: |||
107 AspLeuPheAlaCysSerGluSerLeuArgAsnGlyAsnGlyLeuGlu 123
    |||::: |||::: |||::: |||
223 .....GCTACAGCTCAGTGAAGCGCTCCCAATTCACCGACAC 260
    |||::: |||::: |||::: |||
123 uAsnAlaSerLeuSerGluPheGluLysAsnLysLysIleSerLeuLeu 140
    |||::: |||::: |||::: |||
261 TACCTCTCAGATTATGCAAGAAAGAAAGAGAGGATCATR..... 303
    |||::: |||::: |||::: |||
140 IsSerSerLysGluLysLeuArgArgGluArgGlyLeuTyrCysCysGlu 156
    ||||| ||||| ||||| ||||| |||||
304 .....GAAAGGCGCTCGGATCGATGATTAATAAGATTATCT 342
    ||||| ||||| ||||| ||||| |||||
157 GluLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAla.. 172
    ||||| ||||| ||||| ||||| |||||
343 GAATTAGAGAGACTAGTGCACACGCTTTGAAACAGAGACTCCAA 392
    ||||| ||||| ||||| ||||| |||||
173 .....AlaSerValLeuGluLarPheValAspTyrValLysTyrI 186
    ||||| ||||| ||||| ||||| |||||
393 GTTAGAAAAGGCTGAATATTCGAAATGACAGTGGATTTGAAAGATG 442
    ||||| ||||| ||||| ||||| |||||
186 LeArgGluLysIleSerProAlaValMet.....AlaGlnIleThr 199
    ||||| ||||| ||||| ||||| |||||
443 TCAGGCTACAGGGGGAAGGCTACTTGTATGCCATGCTGTGCACA 492
    ||||| ||||| ||||| ||||| |||||
200 GluAlaLeuGlnSerAsnMetArgPheCys.....LysLysGlu 212
    ||||| ||||| ||||| ||||| |||||
493 GACTTCATGAGCATGTGATCCAGAGAGTCTGCACAGAAAGTGGCTGTA 542
    ||||| ||||| ||||| ||||| |||||
212 nglnthProIIeGlulLeuSerLeuProGlyThrValMetAlaGlnArg 229
    ||||| ||||| ||||| ||||| |||||
543 CCAAGCTCAGTGAAGGCGCTTGACCCGTGAGACCCACATACCGCTGCC 592

```

```

229 luAsnSerValMetSerThrTyrSerProGluArgGlyLeuGlnPheLeu 245
    |||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
593 TTGCTCTCATCTCAGCACTGCTGCTCCACGGG..... 627
246 ThrAsnThrCysTrpAsnGlyCysSerThrProAspAlaGluSerSerLe 262
    ..... 627
262 uAspGluAlaValAlaGValProSerSerSerAlaSerGluAsnAlaIleG 279
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
628 ...GAGCAGCAGTGTATGACATCTCCATGCGCCACACCACTCACCCT 673
279 LysAspProGlyLys.....ThrHisIleSerSerAlaAla 290
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
674 TGCACCTTCACACCTGCGCAGCTGCTTCCACCATCTCCACCACTCCG 723
291 LeuSerLeuAsnSerLeuHisThrValAlaGlyTyrSerLysValThrPr 307
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
724 CTCAGGCCCAATGAGACTCCACACA.....TCAGAGTCAACCC 761
307 oSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuP 324
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
762 ATGTGCGCTATCCACATCT..... 780
324 roSerAlaMetProProValSerSerPheSerLeuGlyThrAlaLeuLeu 340
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
781 ..TCGAGAGTCTCTCTCTCATGCTGCTGCTCTCTCTCATCAGCAACGTTT 828
341 GlyTrpAlaArgArgAlaLeuHisIleProThr 351
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
829 GCCCATGCAGATTCGTCTCTCGATGCCATCA 861

```

seq_name: gb-ro:AF173902

seq_documentation_block:

LOCUS AF173902 2541 bp mRNA ROD 28-FEB-2000

DEFINITION Mus musculus basic helix-loop-helix factor 1 (CHP1) mRNA, complete cds.

ACCESSION AF173902

VERSION AF173902.1 GI:6636410

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 2541) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Chin,M.T., Maemura,K., Fukumoto,S., Jain,M.K., Layne,M.D.,

Watanabe,M., Hsieh,C.M. and Lee,M.E.

Cardiovascular basic helix loop helix factor 1, a novel

transcriptional repressor expressed preferentially in the

developing and adult cardiovascular system

J. Biol. Chem. 275 (9), 6381-6387 (2000)

JOURNAL MEDLINE 20158950

AUTHORS 2 (bases 1 to 2541)

TITLE Chin,M.T. and Lee,M.-E.

JOURNAL Direct Submission

Submitted (30-JUL-1999) Cardiovascular Division, Brigham & Women's

Hospital, 75 Francis Street, Boston, MA 02115, USA

FEATURES

source: 1.2541

/organism="Mus musculus"

/db_xref="taxon:10090"

/dev_stage="embryo"

1.2541

/gene="CHP1"

128.1147

/gene="CHP1"

/codon_start=1

/product="basic helix-loop-helix factor 1"

/protein_id="AF20174.1"

/db_xref="GI:6636411"

/translation="MKRPCEETSTESDLDETIDVGSENNYPGHATSSVSRNSPTTTS

```

OIMARRKRRIIEKRRRIINSLSELRVLPFAFEKOSAKLEKEILQMTVDHLM
LOATGKGKGFDAHALATDPMISGFRECLTEVARYISVGLDPSPLRVRLVSHLSTC
ASQREAAVYVTSMAHHNHPHHPHNAAPFHHPTLLDLPNGHSTSETPCRSTSEV
PSAHSALITATFADSDALRMPSGGTVAFCVPPJSTLSLSATVHAANAATAAHH
SEPLSFEGAFPMPLPSNAAAAAANAATAISPLSVSAASSPQOTSTGINPKYOPWGT
EVGAF"
BASE COUNT      719 a      604 c      551 g      667 t
ORIGIN

```

alignment_scores:

Quality	Ratio	Length	Gaps
114.50	0.774	311	10

Percent Similarity: 47.588 Percent Identity: 21.543

alignment_block:

US-09-389-000-2 x AF173902 ..

```

Align seg 1/1 to: AF173902 from: 1 to: 2541
57 ProLeuThrMetGluLysMetSerAsnValValLysTyrTrpThrCys 73
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
92 CCGGTAGCGCAGTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 141
73 sProSerAsnThrValLysThrGluAsnAlaThrGlyProGluLysLeu 90
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
142 TGAGGAACGACCTCCGAAACGACCTG.....GAGCAGACCATCG 182
90 LysLeuProLeuGlnArgSerTyrSerGluHisLeuGlyTyrPheProThr 106
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
183 ACGTGGGAGGAGCAGACAAATATACCTGGGC..... 214
107 AspLeuPheAlaCysSerGluSerLeuArgAsnGlyLysGlnGlyLeu 123
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
215 .....GCTACAGCTCAGTAGTAGGTCAATTCACCGAC 252
123 uAsnAlaSerLeuSerGluPheGluLysAsnLysLysIleSerLeuLeu 140
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
253 TACCTCTCAGATTATGCGCAAGAAAGAAAGAGGATCATATA..... 295
140 sSerSerLysGluLysLeuArgArgGluArgIleLysTyrCysCysGlu 156
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
296 .....GAAAAAGCGCTCGGATCGAATAAATAACAGTTTATCT 334
157 GluLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAla.. 172
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
335 GAATTGAGAGACTAGTGCACAGCTTTGAAAAACAGAGATCTGCCAA 384
173 .....AlaSerValLeuGluAlaThrValAspTyrValLysTyrI 186
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
385 GTTGAAGAAAGCTGAATATTTGCAGAAATGACAGATGATTTGAAGATGC 434
186 LeuArgGluLysIleSerProAlaValMet.....AlaGlnIleThr 199
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
435 TCCAGGCTACAGGGGTAAAGGCTACTTGTATGCCCATCTCTTGCACA 484
200 GluAlaLeuGluSerAsnMetArgPheCys.....LysLysGlu 212
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
485 GACTCTATGAGCATGTGATTCGAGAGTGTGACAGACAGTGGCTAGTA 534
212 ngAlnThrProIleGluLeuSerLeuProGlyThrValMetAlaGlnArg 229
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
535 CCTAAGCTCAGTGAAGCCTTGACCGCTGAGCCCACTACGGGTGGCC 584
229 luAsnSerValMetSerThrTyrSerProGluArgGlyLeuGlnPheLeu 245
    |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
585 TTGCTCTCATCTCAGCAGCCTGTGCTCCACGGG..... 619
246 ThrAsnThrCysTrpAsnGlyCysSerThrProAspAlaGluSerSerLe 262
    ..... 619
619 ..... 619
262 uAspGluAlaValAlaGValProSerSerSerAlaSerGluAsnAlaIleG 279

```



```
|||||: : ||||| : : : : :
620 ...GAGGAGAGATGATGATCTCCAGGGCCACACCATCACCCCT 665
279 LysProTyrLys.....ThrHisSerSerAlaAla 290
      |||: : |||: :
666 TGCACCTTCACACTGGGAGCTGCTTCCACCATCTCCCCACAGCCCTG 715
291 LeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThrPr 307
      |||: : |||: : |||: : |||: :
716 CTCAGGCCCAATGACTCCACACA.....TCAGAGTCAACCCC 753
307 oSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuP 324
      | : : : : :
754 ATGTCCCTATCCACATCT..... 772
324 roSerAlaMetProPheValSerSerPheSerLeuGlyThrAlaLeuLeu 340
      ||| : : ||| : : ||| : : |||: :
773 ..TCAGAGTGCCTTCTGCTCATGGCTGCTCTCTCTCCACAGCAGCTT 820
341 GlyTrpAlaArgArgAlaLeuHisIleProThr 351
      : : : ||| : : |||: : |||: :
821 GCCCATGCAGATCTGCTCTTCGATGCCATCA 853
```

THIS PAGE BLANK (USPTO)

OM of: US-09-389-000-2 to: N_Geneseq_36.* out_format : pfs

Date: Jan 19, 2001 5:12 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frim+ -p2n.model -DEV=xlp  
-O=/cgn2_1/uspro.spool/US09389000/runat_17012001_152619_25058/app_query.fasta_1.467  
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM_ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09389000 -CGN1_1.133 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-389-000-2  
Query length: 405  
Database: N_Geneseq_36.*  
Database sequences: 480022  
Database length: 187831343  
Search time (sec): 88.730000
```

score_list:

Sequence	Strid Orig	ZScore	Escore	Len	Documentation
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:294275		2095.00	3660.08	4.4e-196	2128
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:257039		115.00	186.33	0.0137	568
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:227531		113.50	164.07	0.2379	3045
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:Q75165		113.00	150.05	1.44	9370
/SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:Q75164		113.00	150.02	1.44	9391
/SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:T11413		109.50	163.80	0.2461	1704
/SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:Q73000		108.50	161.03	0.3510	1857
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T70134		108.00	175.25	0.0567	510
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T70133		108.00	174.65	0.0612	537
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T70132		107.00	165.36	0.2016	1023
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T70130		105.00	176.56	0.0479	290
/SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:Q38720		103.00	146.27	2.33	2866
/SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V27865		103.00	134.45	10.61	2948
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X18940		102.00	144.18	3.05	2948
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T62360		101.50	151.69	1.16	1437
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X33679		101.50	143.44	3.35	2911
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:X33679		101.50	140.32	5.00	3804
/SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:Q21097		101.00	150.11	1.42	1526
/SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V24018		101.00	147.90	1.89	1844
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T73287		100.00	139.64	5.45	3214
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T62360		100.00	139.64	5.45	3214
/SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V40019		99.50	141.79	4.14	2481
/SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V40019		99.50	141.79	4.14	2481
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X26611		99.50	138.88	6.02	3183
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X13132		98.00	116.80	102.15	16788
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:V72220		97.50	140.32	5.00	2080
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:A34739		97.50	140.21	5.07	2100
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:A34751		97.50	131.37	15.76	2475
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:A34741		97.50	129.37	20.37	5312
/SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:Q58522		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X89220		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X88221		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:Z39404		97.50	129.27	20.62	5355
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:Z39404		97.50	129.27	20.62	5355
/SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:Q75977		97.50	127.65	25.38	6151
/SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T46184		96.00	139.99	5.22	1707
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X91990		95.50	130.39	17.87	3600

/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X20249	-	95.50	89.79	3.3e+03	116277
/SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:T08974	+	95.00	128.53	22.69	3915
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X15635	+	95.00	128.53	22.69	3915
/SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:A12989	-	94.50	142.89	3.60	1062
/SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X58669	+	94.50	137.98	6.75	1617

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:294275

seq_documentation_block:

ID 294275 standard; cDNA; 2128 BP.

AC 294275;

DT 03-JUL-2000 (first entry)

DE Human PHELIx CDNA clone GPC1C12.

KW PHELIx; human; testis-specific; transcription factor;

KW prostate cancer; bladder cancer; ovary cancer; testicular cancer;

KW gene therapy; diagnosis; vaccine; chromosome 13q13.1-13.3; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 733..1953

FT /tag= a

PN WO200012709-A2.

PD 09-MAR-2000.

PF 31-AUG-1999; 99WO-US20137.

PR 31-AUG-1998; 98US-0098610.

PT 31-OCT-1998; 98US-0106524.

PT WPI: 2000-237872/20.

PT P-PSDB: Y79269.

PT Testis specific Helix Loop Helix proteins expressed in cancers and

PT useful for the prevention, diagnosis and treatment of prostate, bladder

PT and ovarian tumors -

PT Claim 5; Fig 2A-D; 62pp; English.

This sequence is that of cDNA encoding human PHELIx (see Y79269),
a novel basic Helix Loop Helix protein thought to act as a
transcription factor. PHELIx normally exhibits a testis-specific
expression pattern but is up-regulated in prostate and other types
of cancer. The cDNA clone, termed GPC1C12 (ATCC 98956), was
isolated from a normal testis cDNA library using a suppression
subtractive hybridization method. The gene maps to chromosome
13q13.1-13.3. The invention provides diagnostic and therapeutic
methods useful in the management of various cancers which express
PHELIx, including prostate cancer, bladder cancer, ovarian cancer
and testicular cancer, including therapies aimed at inhibition the
transcription, translation, processing or function of PHELIx. The
expression pattern of PHELIx suggests that is an ideal target for a
cancer vaccine approach to prostate cancer.

Sequence 2128 BP; 675 A; 463 C; 464 G; 526 T; 0 other:

alignment_scores: Quality: 2095.00

Ratio: 5.173

Length: 405

Gaps: 0


```

1889  :|::: ||| ||| |::: |:::
1890  AGACACATGCTACACGATCTTACAGCATGTATCAAACTTGCTGGGTATT 1938
310  AspaIatrrhAlaValrhrAsnGlnAsnIleSerIleHisLeuProSerAl 326
    ||| |::: |::: |::: |:::
1939  GATGAAAGATGACCTACTGCTGATGATACCAAGCTGCTGATCTACTGAGA 1988
326  aMetProProVal 330
    ||| |::: |::: |::: |:::
1989  AATGCCACCCTT 2001

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:Q75165
seq_documentation_block:
ID      Q75165 standard; cDNA; 9370 BP.
XX      Q75165;
XX      10-AUG-1995 (first entry)
XX      DT
XX      AF-4 transcript (cDNA clone kcl 6).
XX      DE
XX      AF-4 lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;
XX      KW chromosomal translocation; abnormality; detection; AF-4 gene;
XX      KW ALL-1/AF-4 chimeric gene; t(4;11) translocation; probe; ds.
XX      OS
XX      Homo sapiens.
XX      Location/Qualifiers
XX      FH Key 469..4032
XX      CD CDS /*tag= a
XX      FT 3'UTR 4033..9370
XX      FT /*tag= b
XX      FT /note= "unusually long 3'-UTR, includes multiple
XX      FT AATAA sequences located 20 nucleotides
XX      FT 5' of the polyA, as well as in several
XX      FT upstream positions; it also contains
XX      FT several stretches of T"
XX      FT
XX      MO9426930-A.
XX      PN
XX      24-NOV-1994.
XX      PD
XX      22-APR-1994; 94WO-US04496.
XX      PF
XX      14-MAY-1993; 93US-0062443.
XX      PR
XX      (UYJE-) UNIV JEFFERSON THOMAS.
XX      PA
XX      Canaanl E, Croce C;
XX      PI
XX      WPI: 1995-006818/01.
XX      DR P-PSDB; R66451.
XX      PT
XX      PT New acute lymphocytic leukaemia gene prods. - used for the
XX      PT diagnosis and treatment of leukaemias, partic. acute
XX      PT lymphoblastic or nonlymphoblastic leukaemia
XX      PS
XX      Claim 2; Page 99-106; 207pp; English.
XX      CC
XX      cDNA clones containing the two reciprocal ALL-1/AF-4 RNA junctions
XX      CC were cloned from RNA of the RSK 11 cell line carrying the t(4;11)
XX      CC chromosome translocation. AF-4 specific probes obtained from the
XX      CC clones were used to screen cDNA libraries prepared from RNAs of t
XX      CC K562 and KCl22 haematopoietic cell lines. Positive clones were
XX      CC sequenced and used as probes. Overlapping clones spanning most o
XX      CC all of the 9.5kb AF-4 transcript were obtained. Comparison of
XX      CC different clones suggested an alternative first exon for AF-4 so
XX      CC that AF-4 codes for at least 2 proteins varying at their termin
XX      CC oligonucleotide probes of at least 15 nucleotides derived from
XX      CC this sequence (Q75165) and able to identify chromosome
XX      CC abnormalities within the AF-4 gene of chromosome 4 are claimed.
XX      CC

```



```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:Q73000
seq_documentation_block:
ID      Q73000 standard: cDNA: 1857 BP.
XX      Q73000:
XX      08-JUN-1995 (first entry)
XX      Petunia ph6 gene.
XX      Vacuolar pH; pH gene; Petunia; ss.
XX      Petunia hybrida strain V26.
XX      Key      Location/Qualifiers
XX      CDS      481..1632
XX      FT      /tag= a
XX      misc-feature 484..1629
XX      FT      /tag= b
XX      FT      /label= Claimed
XX      WP09423561-A.
XX      27-OCT-1994.
XX      15-APR-1994: 94WO-US04173.
XX      PR      16-APR-1993: 93US-0049282.
XX      (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX      Chuck GS, Courtney-guterson N, Dooner HK, Keller J;
XX      MJjar CS, Ralston EJ;
XX      WPI: 1994-341349/42.
XX      DR      P-PSDB: R62656.
XX      Petunia Ph gene and constructs containing it - for alteration of
XX      vacuolar pH used in the formation of blue flowers
XX      Claim 7: Page 42-44; 62pp; English.
XX      The V26 strain of Petunia was used. Poly A RNA was isolated from
XX      total RNA from flower buds and used to generate a cDNA library in
XX      the vector lambda ZapII (Stratagene). The SstI to BamHI fragment at
XX      the left hand side of Ac was used to isolate pPet14-1 (contg. ph6
XX      cDNA), the sequence for which is claimed (see Q73000 FT). The AA
XX      sequence comprises a helix-loop-helix structural motif starting with
XX      the sequence NHVLAEK (starting at residue 192) and extending to the
XX      sequence KKKVQDE (ending at residue 245). Proteins including this
XX      motif include the myc family of oncogenes, regulators of neuron and
XX      muscle development, and regulators of segmentation and organ
XX      patterning in Drosophila. In plants, the motif is found in R(S),
XX      a protein involved with regulating anthocyanin synthesis in Maize.
XX      Comparison of the ph6 and R(S) sequences over the 54 AA helix-loop-
XX      helix region detected 57% AA identity. Outside this region only 18%
XX      identity was found.
XX      Sequence 1857 BP; 565 A; 415 C; 413 G; 464 T; 0 other;
alignment_scores:
      Quality: 108.50      Length: 306
      Ratio: 0.714      Gaps: 15
Percent Similarity: 49.673      Percent Identity: 23.559

```

```

Alignment block:
US-09-389-000-2 x Q73000 ..

Align seg 1/1 to: Q73000 from: 1 to: 1857

6 ValProSerSerLeuSeraIaGlLueGluAlaIleuSleuIleAr 22
:::|||||::: ||| ::|||::|
670 TITCAAGGGCGCCCTAGTATGAATTA..... 699
22 gpehgLyIsLysAsnThrHis..SerLeuPheValPheIleIleP 38
:::|||||::: ||| ::|||::|
700 .....TCCAAAGAAGACACACTTCTCATAAACAGTGCGACCATTC 742
38 rogluAnpHelys..... Gly 43
|||::|::|::|
743 TTGAACAACCCTTAACCAAAAGCTCCAATTTCCTTACCATATGAGGC 792
44 CYSILeserLyHisgIuMetaspIIeAlaIleuThrgLuPLeuThrHe 60
|||||::|::|
793 TGATATTCC..... 801
60 tGIuLYMeSeRasNvaIalYstYrrPhThrCYSProSeRaNsnt 77
:::|||||::: ||| ::|||::|
802 .CAAAACAACCATTCTGCCCTTCCAAAGGTGCCAGC...CCCAGCACCA 847
77 hrvalYstThrgIuaAnaIaThGlyPProGLuGLuLeugLYLeuProLeu 93
|||||::: ||| ::|||::|
848 CCCTCTCCACCCATTCTTTGACGGGGGGCCACCTCCGGCCAGTGcTG 897
94 GlnArySerTyriSerGluHisLeugLYTYrPheProThraspleuPeAl 110
:::||||::: ||| ::|||::|
898 CTCAAAGACATACTATTCTGTCCATTCTTCACACTTAATACCAAC 947
110 acYSseRGlu..SerLeuArgasnGLyasnGLyLeuGLuEuaSnlaS 126
TGCAAGTGAAGTTTCTCCAAAGTCAGCTGACGCTACCACTGTGATTCTCT 948
126 er..LeuSerGluPheGLuLYasnlYsSYsile..... 136
||| ::|||::|::|::|::|::|::|
998 CCACCTCATCTGCCCTTTCGAAAAGGGTGATGATNAACAAGAAGAGCCT 1047
137 SerLeuLueHisSerSerLYsgLUlysLeuArgargLUarGLilelYsTY 153
||| ::|||::|::|::|::|::|::|
1048 AGTGAAGAACCATGACTTCTGCTGAACGACGCGCGTAGAAGAAAAAGCTCAACGA 1097
153 rCySGySLuGLuLeuThrgThrHeuLeuProTYrValYsGLyArgLYsa 170
1098 ACGGTTATTCATTATTAAGGTCACTTGCTTGTCTTTGTT..... ACGAAA 1141
170 snASpaIIaIaSeRvAlleuGLuAlaIthrValaSPryValLYsTYrIle 186
||| ::|||::|::|::|::|::|::|
1142 TGGATTAAGGCTCATCTTCGTGGACACCACTAGATAATGTCAACAGAGTTA 1191
187 ArgGLuLYsIleserPro..AlaValMetalaGLinleThrgLualale 202
|||::|::|::|::|::|::|::|::|
1192 CGTAAAGAAAGTTGAGATCTTGAAGCTAGAGCCAAATGACAGCGAGGCT.. 1239
202 uGINserAsnMeArGpheCYsLYsGLynGLInThPrOleGLeuS 219
1240 ..... ACGGTGACAGCAA 1252
219 eRIeUpRoGLYthrValMetalaGLINargLUasnSerValmetSerThr 235
:::|||||::: ||| ::|||::|
1253 AGGATACAGGTAAGTGAAGCTTCCAAAGAAAGCGCT..... 1290
236 TyrSerProGLuArgLYleuGLInPheLeuThraSnThrcYSTrpasngI 252
:::||||::: ||| ::|||::|
1291 .....AAGGAGCAATGAAGAAATAGTGAAGAAAGTGTGGTGAGCG 1331
252 Y..... CySerThrProASpaIIaGLuSerSerLeuLuspg 264
:::|||||::: ||| ::|||::|
1332 ACAAGCAAAAGATACGGCACTTTCGCCCTCAACAGACATGA....GAGG 1378

```



```

100 CGGGCCCGCCAGGCCCCCTACGGTGGCCGAGGTTCCAGC..... 140
136 lIeSerLeuHnIseSer.....LysGluLysLeuArgGluAr 150
141 ...TCTCGCTTCGGCGGAAAGCTGATCAGCAGAGGAGGACGAGAAAGCG 187
150 gIleLysTyrCysGluGlnLeuArgThrLeuLeuProTyrValLysG 167
188 GATGTGCTTGAGCTGTGAGCTGTGCGGCGCTGCTGCCAGTTGATG 237
167 lArGlysAsnAspAlaAlaSerValLeuGluAlaThrVal 180
238 GCCGGCGGAGGACATGCGCTGCTGCTGAGATGTCTGTT 278

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:Q38720

seq_documentation_block:

ID Q38720 standard; DNA; 537 BP.

AC Q38720;

DT 15-JUL-1993 (first entry)

DE Max cDNA.

KW Med; max. myc; c-myc; helix-loop-helix zipper; leucine zipper;
KW helix-turn-helix; diagnosis; prognosis; cancer; malignancy;
KW neoplasm; tumour; studying embryogenesis; study gene regulation.

FT Key Location/Qualifiers

FT CDS 28..508

XX W09305056-A.

XX 18-MAR-1993.

XX 09-SEP-1992; 92WC-US07629.

XX 09-SEP-1991; 91US-0756195.

XX 23-JUN-1992; 92US-0903710.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Ayer DE, Blackwood EM, Eisenman R;

XX WPI; 1993-100913/12.

XX P-PSDB; R33386.

XX Helix-loop-helix zipper protein named Max - associates with MYC
PT or Mad polypeptide(s) useful as diagnostic or prognostic tools
PT for diverse types of cancer

XX Claim 15; Fig 2; 93p; English.

XX This sequence represents cDNA for Max protein. A random primed
CC lambda gt11 expression library was derived from a baboon lymphoid
CC cell line 3545 as described in [RL Idzerda et al. Proc. Natl. Acad.
CC Sci. USA. 86, 4659 (1989)]. Phage from this library produce nearly
CC full length beta-galactosidase proteins fused with the ORF of the
CC directionally cloned cDNAs. More than 10power4 plaques were
CC screened for their ability to interact with 125-I labeled
CC GST-Myc92. Several positive plaques were identified, and Max 11
CC and 14 survived multiple rounds of plaque purification. Sequence
CC analysis of these two clones together with those derived from a
CC Max lambda gt10 library enabled the sequence shown below to be
CC deduced.

XX Sequence 537 BP; 165 A; 158 C; 133 G; 81 T; 0 other;

alignment_scores:

Quality: 103.00 Length: 136
Ratio: 1.272 Gaps: 3
Percent Similarity: 59.559 Percent Identity: 25.735
Alignment_block:
US-09-389-000-2 x Q38720 ..

Align seg 1/1 to: Q38720 from: 1 to: 537

```

140 HisSerSerLysGluLysLeuArgArgGluArgGileLysTyrCysGle 156
109 CATATGCACTGGAACGAAAGAGTGGAGGACCAATCAAAAGACAGCTTCA 158
156 uGlnLeuArgThrLeuLeuProTyrValLysGluArgLysAsnAspAla 173
159 CAGTTGGCGGACTGATGCCATCTCACTCAAGGAGAGAGGATCCAGG 208
173 lAserValLeuGluAlaThrValAspTyrValLysTyrIleArgGluLys 189
209 CCCAATCTTAGACAAAGCCACAGAGTATATCTACTATATCGAAGGAA 258
190 lIeSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
259 .....AACCAACACACAGCAAGATAT 281
206 tArgPheCysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
282 TGACAGCTCTCAAGCGGCAATGCTTCTTGAGCAGACGACTCGTGAC 331
223 hrValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerProGlu 239
332 TGAGAAAGCG...AGTTCAGTGGCCCAATCCAGACCACTACCCCTCC 378
240 ArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerThrPr 256
379 TCAGACAACAGCTCTACACCAAC..... 402
256 oAspAlaGluSerSerLeuAspGluAlaValAlaArgValProSerSerSera 273
403 ...GCCAAGGCGACACCATCTCTGCTTTGATGGGGCTCAGACTCCA 448
273 lAserGlu 275
449 GCTCAGAG 456

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T43383

seq_documentation_block:

ID T43383 standard; cDNA; 2866 BP.

XX T43383;

XX 11-MAR-1997 (first entry)

XX Human cytokine response gene CR8.

XX Cytokine response gene; CR8; interleukin-2; IL-2;

XX ligand-stimulated gene expression; diagnosis; therapy;

XX transcription factor; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 240..1478

XX W09639427-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-US09194.

XX 05-JUN-1995; 95US-0465585.

```

PR 05-JUN-1995; 95US-0461379.
PR 05-JUN-1995; 95US-0462337.
PR 05-JUN-1995; 95US-0462390.
PR 05-JUN-1995; 95US-0463074.
PR 05-JUN-1995; 95US-0463081.
XX
XX (DART-) DARTMOUTH COLLEGE.
XX
XX Beadling C, Smith KA;
XX
XX MPI: 1997-043062/04.
XX
XX P-PSDB; W08140.
XX
XX Cytokine response proteins and genes - used in the detection and
XX therapy of diseases caused by a mutation in the CR coding region
XX
XX PS Disclosure: Page 36-40; 81pp; English.
XX
XX 8 clones (T43376-83) contg. interleukin-2 (IL-2)-induced genes
XX were isolated from a human IL2 receptor-positive T blast cell cDNA
XX library following IL-2 stimulation. 6 Of these ligand-induced genes
XX (CR1, 2, 3, 5, 6, 8) are novel. The CR8 gene encodes a 45 kDa
XX protein (W08140) that is a putative DNA binding protein and
XX transcription factor. CR genes and polypeptides (W08133-40) are
XX useful as diagnostic or therapeutic agents; CR gene sequences can
XX be used to detect and treat allelic mutations.
XX
XX SO Sequence 2866 BP; 703 A; 751 C; 716 G; 654 T; 42 other;

alignment_scores:
  Quality: 103.00      Length: 267
  Ratio: 0.769        Gaps: 19
  Percent Similarity: 50.187  Percent Identity: 28.464

alignment_block:
US-09-389-000-2 x T43383  ..

Align seg 1/1 to: T43383 from: 1 to: 2866

74 ProSerAsnThrValIleuGluSerGluAsnAlaThrGlyProGluGluLeuG1 90
   |||:::.....:|||||:::|||||:::
233 CCCCAGCATGAGGAGATCCCGCCGCGCAACCCGCCG..... 274
   90 yLeuProLeuGlnArgSerTyrSer..GluHisLeu.....GlyTyr 103
   |||||:::|||||:::|||||:::|||||
275 .CTGCGTCCCAAGCAAGCGGACTGTGACGACCGACCTACACGCGATG 323
   104 PheProThrAspLeuPheAlaCysSerGluSerLeuArgAsnG1ysaG1 120
   ::|||:::|||||:::|||||:::|||||
324 TACCCTGCCCAATGTACCAAGTGTACCAAGTACAGACG.....GG 364
   120 yLeuGluLeuAsnAlaSerLeuSerGluPheGluAsnLysLysLies 137
   |||:::|||||:::|||||:::|||||
365 AATAAAGCGAGGAGGACGACGACGACCTACAAA..... 401
   137 erLeuLeuHisSerSerLysGluLysLeuArgArgGluArgLysTyr 153
   |||:::|||||:::|||||:::|||||
402 ..TTGCCGACCGGCTCTTCGAGAAAAAGACGTGACCGGATTAACGAG 449
   154 CysCysGluGlnLeuArgThrLeuLeuPro...TyrValLys..... 166
   |||:::|||||:::|||||:::|||||
450 TGCATGCCCGCAGCTGAAGATCTCTACCCGACATCTCAAAACTTACAC 499
   167 ....GlyArgLysAsnAspAlaIaIaSerValLeuGluAlaThrValAsp 182
   |||:::|||||:::|||||:::|||||
500 TTGGGGTCACTTG...GAAAAAGCATGGTCTTGAACCTTACCTTGAAC 546
   182 yValLysTyrIleArgGluLysLiesSerProAlaValMetalGlnIle 198
   ::|||:::|||||:::|||||:::|||||
547 ATGTGAAGACACTACAAACCTAATTGATG.....CACACACAGAAA 590
   199 ThrGluAlaLeuGlnSerAsnMetArg..... 207

```

```

591 ATCATTTGCCCTGCAGAGTGTGTACAACTGCTGACGCTCTCAGCGAATA 640
208 .....PheCysLysLysGlnGlnThrProIleG 217
641 TGTGAAACAGCTCAAGAGATGTTCTGCTCAGGTTTCCACAGCATGTGCC 690
217 LuLeuSerLeuProGlyThrValMetalGlnArgLysAsnSer..... 734
691 GGGAGGTGCTT.....CAGTATCTGGCCAGCAGCAGACACTCGGGAC 744
232 ...ValMetSer..ThrTyrSerProGluArgGlyLeuGlnPheLeuThra 247
735 CTGAAGCTTTCGCGAGCTGTGCACC.....ACCT 763
247 snThrCysTrpAsn..GlyCysSerThrProAspAlaGluSerLeuAs 263
764 CCACCGGCTGCTCGAGCTGCT.....GCAAGGTGATCCTCCA 804
263 pGluAlaValAlaArgValProSerSerSerAlaSerGluAsnAlaIleGly 280
805 GGAAGCATCAGA...CCAGCTCCCAAGTGTAT.....GGAC 839
280 sProTyrLysThrHisLiesSerSerAlaAlaLeuSerLeuAsnSerLeu 296
840 TTCAGGAAACACCAGCTCTCCGGCCAAAGTTCGGA..... 877
297 HisThrValArgTyrTyrSerLysValThrProSerTyrAspAla 311
878 .....AGTCTCTGGGAAAAAAGCTGCTGCCAGCATCATCCACGG 913

seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:V27865
seq_documentation_block:
ID V27865 standard; cDNA; 7881 BP.
XX
XX V27865:
XX
XX 12-OCT-1998 (first entry)
XX
XX Human telomerase RNA interactive protein-1 (TRIP1) cDNA.
XX
XX TRIP1; human; telomerase RNA interactive protein-1; cancer; AIDS;
XX ageing; therapy; ss.
XX
XX Homo sapiens.
XX
XX MO9821343-A1.
XX
XX 22-MAY-1998.
XX
XX 13-NOV-1997; 97WO-US21248.
XX
XX 16-OCT-1997; 97US-0951733.
XX 15-NOV-1996; 96US-0871189.
XX 11-JUN-1997; 97US-0875039.
XX
XX (AMGE-) AMGEN CANADA INC.
XX (AMGE-) AMGEN INC.
XX
XX Harrington LA, Robinson MO:
XX
XX MPI: 1998-297946/26.
XX P-PSDB; W61347.
XX
XX New nucleic acid encoding human telomerase protein-2 - used for
XX regulating telomerase activity, e.g. for treating cancer or acquired
XX immune deficiency syndrome
XX
XX Example 2; Fig 1; 150pp; English.
XX
XX This cDNA sequence codes for human telomerase RNA interactive
XX protein-1 (TRIP1, see W61347). It was obtained by screening a

```



```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TUV-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 469..4032
US-08-545-860D-27

alignment_scores:
Quality: 113.00 Length: 424
Ratio: 0.589 Gaps: 22
Percent similarity: 45.283 Percent identity: 21.462

alignment_block:
US-09-389-000-2 x US-08-545-860D-27 ..
Align seg 1/1 to: US-08-545-860D-27 from: 1 to: 9370
25 LysLysLysAsnThrHisSer LeuPheValPheIleIleProGluAsn 41
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2888 AGAAGAAATCAATCAACAGTCATCTTCACTTCACTCCCAAGAA 2937
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2938 TCTTCTAAACAAA.....GCCCTC 2957
55 ThrGluProLeuThrMetGluLysMetSerAsn.....ValVal. 67
2958 CAGGCCCTCTCAAGCTCTCAAGGAAGAAAGTCTCCCCCGCCACCG 3007
66 .....ValVal. 67
3008 TCTCCTGCTCTCCCAAGAGCCAGCCAGCCCTGCACTTAAGAGGTCAAG 3057
68 LysTyrTrpThrCysProSerAsnThrValLysThrGluAsnAlaThr 84
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3058 CCGAGAGCAGACACCTGTGGCCAGACCTCCCAAAAGTGCAGCAGTAC 3107
84 r.....GlyProGluLeuGlyLeuProLeuGln..... 94
3108 CAAGAGCAACACCAAGACTTTCATCTCCCAAGCAGAGAGAGTAGAGC 3157
95 .....ArgSerTyrSerGluHisLeuGlyTyr..... 103
3158 GGAAGGGCTCCAGAAAGCTCTCGAGACCAAGAGGTCTCCGAGAGTACT 3207
```

```

104 .....PheProThrAspLeuPheAlaCysSerSerGluSerLeuArgAs 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3208 GCAAAATCCTTTCCAGTG.....CCTTTGTCACAA 3239
117 ngLysnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGlyLysAsnL 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3240 TGGTACTCT.....AAACGAGGAGAGCCTCAAGTGAAGTTTGACA 3280
134 ySLysIleSerLeuLeuHis.....SerSerLys 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3281 AACAAACAGACAGACCTTCACATGAGGAGCGCAAAAGATGAGCAGAAA 3330
144 GluLysLeuArgGluArgGlyLeuLysTyrCysCysGluGlnLeuArgTh 160
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3331 GCAGAGTTAATGACGCGACAGGAGGTGGAAGGCTTTAAGTACTCGAAGC 3380
160 rLeuLeuProTyrValLys.....GlyA 168
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3381 CGCTGTGCTTCATTTAGTGCGGATTCGACAGAGCTGAAAGCCAGT 3430
168 rGlyAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValLys 184
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3431 CATCCAGTCAGACTTACTGCTGCTCTCAGAAACGTGATGATCTATAA 3480
185 TyrIle.....ArgGly 188
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3481 TTCATATATGTCATTAATTCCTTCAGATGCCACAGCGCCACAAAGA 3530
188 uLysIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerA 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3531 GAATAATATTGCTGTTTATGATCGTGCCAGTCCATTGGAACATGG 3580
205 smMetArgPheCysLysGlnGlnIleThrProIleGluLeuSerLeuPro 221
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3581 CGATGCTTCGTTGTAATAAAGC.....ATGCAATAAAGATATCT 3621
222 GlyThrValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerPr 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3622 CCGTACTCTTAATTAACACTGCGAGAGTTCTCCAAAGTCCGCGAGCAC 3671
238 oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerT 255
3672 T.....TCTCATGCAATTGCAAGCACAGGCA 3697
255 hrProAspAlaGluSerSerLeuAspGluAlaValArgValProSerSer 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3698 CACCAATCCCTCTTCCCAATGCTCTCGCCAGCTCCGTAAGGATCC 3747
272 SerAlaSerGluAsnAlaIleGlyAspProTyrLysThrHisIleSerSe 288
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3748 CAGTCAAGTCTGCGCAGTGTGGGAGC.....AGTGG 3779
288 rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrSerLysSV 305
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3780 GGTGGCTGCACATATCAGACCCAGTCACATCCAG.....ATA 3820
305 alThrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIle 321
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3821 TGACATCTTCTAT.....GTCAACATC 3843
322 .HisLeuProSerAlaMetProProVal..... 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3844 ACAT...CCCATGTTCTTACCGCTTTGACCTTTGGAGACGGCCGAGC 3890
331 ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTyrAlaArgArgAla 346
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3891 CCTCAGAGAGAAATTAAGATTTCTTGTCTGGCTCAGCAAAATGTGT 3940
347 LeuHisIlePro.....ThrValCysAsnSerPheGly 357
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3941 GCACCTTGCCCTCAACAGACAGTTTGGTGAACSTGTGCACATATACAGA 3990
357 yArgIleLysSerThr 362
```

||||: |||:|
3991 CAGGCTTTTCAGCAGC 4006

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-04496-27

seq_documentation_block:

Sequence 27, Application PC/TUS9404496

GENERAL INFORMATION:

APPLICANT: Croce, Carlo

APPLICANT: Canaan, Eli

TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

ADDRESSEE: Norris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04496

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TUV-1242

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 9370 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 469..4032

PCT-US94-04496-27

alignment_scores:

Quality: 113.00 Length: 424

Ratio: 0.589 Gaps: 22

Percent Similarity: 45.283 Percent Identity: 21.462

alignment_block:

US-09-389-000-2 x PCT-US94-04496-27 ..

Align seg 1/1 to: PCT-US94-04496-27 from: 1 to: 9370

25 LysylslnsanthrhiserLeupheValPheileProgluasp 41

2888 AGAAGCAAAATCAATCATCTTCACTTCATCTCCCAAGAA 2937

41 he.....LysglyCysIleSercIyHISglYmetAspIleValLeu 54

2938 TCTTCAAAACAAA.....GCCCTC 2957

55 ThrGluProLeuThrMetGluYsMetSerAsn..... 65

2958 CAGGCCCTCTCTACAGTCTCAAAAGAAAGTCTCCCCGCCACCG 3007

66ValVal. 67

3008 TGTCTCTGCTCTCCAGAACCCAGACGCTTCAATTAAGAGTCAAG 3057

68 LysTyrTrpThrCysProSerAsnThrValLysThrGluAsnIaTh 84

3058 CGGGAAGCAACACCTGTGGCAGACCCCTCCCAAGTCCAGCATAC 3107

84 r.....GlyProGluGluLeuGlyLeuProLeuGln..... 94

3108 CAAAGCAACCAACAAAGACTTTCATTCCTCCAGACAGAGAGTAGAGG 3157

95ArgSerTyrSergIuHISleuGlyTyr..... 103

3158 GGAAGGCTCCACAACTCTCTCGAGCACAAGGCTTCTCCGAGATAC 3207

104PheProThrAspLeuPheAlaCysSergIuSerLeuArgAs 117

3208 GCAAACTCTTTCACAGT.....CCTTCTTGGCCAAA 3239

117 ngIysangIyLeuGluLeuAsnAlaSerLeuSergIuPheGluYsAsn 134

3240 TGGTAACCT.....AACCAAGGAACCTCAAGTAGATTGAC 3280

134 yslYsIleSergLeuLeuHIS.....SerSergLys 143

3281 AACACAAAGCAGACCTTCATGAGGAGGCAAAAAGATGAACGAGAA 3330

144 GluYsLeuArgArgGluArgIleYsTyrCysGlyGluLeuArgTh 160

3331 GCAGAGTTAATGACGACAGAGGTTGGAAGGCTTTAAGTACCTGGAAG 3380

160 rLeuLeuProTyrValLys.....GlyA 168

3381 GGTCTTCTCTTATTGAGGCGGAATTGCCACAGCTGAAGCCAGT 3430

168 rglYsAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValLys 184

3431 CATCAAGTCAGCTTACTGCTGCTACTCAGAAACTGATTCATTA 3480

185 TyrIle.....ArgG 188

3481 TTCATATGTCATTAAATCTTCTCAGATGCCACAGCCGCAACAGAA 3530

188 ulYsIleSerProAlaValMetAlaGlnIleThrGluAlaLeuSera 205

3531 GAAATATTTTGTCTTTTATGATGCTGCTGCTACTCAGAACTGATTCATTA 3580

205 snMetArgPheCysLysLysGlnGlnThrProIleGluLeuSergLeuPro 221

3581 CGATGTTTGTCTTAAAGAAC.....ATACCATTAAGTATTC 3621

222 GlyThrValMetAlaGlnArgGluAsnSergValMetSergThrTyrSerg 238

3622 CGTACTCTTAAATCACTTCGAGAGTTCTTCCAAAGTGCCGACGACAC 3671

238 ogIuArgGlyLeuGlnPheLeuThrAsnThrCysTyrSergIuSerg 255

3672 T.....TCTCATGTCATTGAAGCAGCA 3697

255 hrProAspAlaGluSergSerLeuAspGluAlaValArgValProSerg 271

3698 CACATACCCCTCTTCCCAATGCTTCTCTCCAGCTCCGAGGCTC 3747

272 SerAlaSerGluAsnAlaIleGlyAspProTyrLysThrHisIleSerg 288

3748 CACTCAAGTCTGTCAGTGGGAGC.....AGTGG 3779

288 rAlaAlaLeuSerLeuAsnSergLeuHisThrValArgTyrSergLys 305

3780 GGTGCTGCATATGACACCCAGCTCAGCATCCAG.....AATA 3820

305 alThrProSergTyrAspAlaThrAlaValThrAsnGlnAsnIleSerg 321

```

      :||| ||||| :|||
3821 TGACATCTCCAT.....GTCACCATC 3843
3322 ..HisLeuProSerAlaMetProIroVal..... 330
3844 ACAT...CCCATGTTCTTACCGCCTTGACCTTGGAGAACGCCCAAGC 3890
331 ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTrpAlaArgArgAla 346
3891 CCTCAGAGAGAGAAATAAAGAAATCTTGTGCTGCGTCAGACAAATGTGT 3940
347 LeuHisIlePro.....ThrValCysAsnSerPheG1 357
3941 GCACCTTGCGCTCAACAGCAGCTTGTGTGACCTGTGCACATATACAGCA 3990
357 ValGlyLeuSerThr 362
3991 CAGGGTTTCAGCAGC 4006

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-320-559-25

seq_documentation_block:
: Sequence 25, Application US/08320559
: Patent No. 5633135
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
: TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESSES:
: ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5633135f1s
: STREET: One Liberty Place - 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/320,559
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/062,443
: FILING DATE: 14 MAY 1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/971,094
: FILING DATE: 30-OCT-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/888,830
: FILING DATE: 27-MAY-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/805,093
: FILING DATE: 11-DEC-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca, Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TTU-0855
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9391 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear

```

```

: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 421..4053
US-08-320-559-25

Alignment scores:
  Quality: 113.00      Length: 424
  Ratio: 0.589
  Percent Similarity: 45.283      Percent Identity: 21.462

Alignment_block:
US-09-389-000-2 x US-08-320-559-25 ..

Align seg 1/1 to: US-08-320-559-25 from: 1 to: 9391

25 LysLysLysAsnThrHisSerLeuPheValPheIleIleProGluAsn 41
: |||||
2909 AGAAGGAAATCAATGACAGCATCTTCACTTCACTCCACAAAGCA 2958
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
: |||
2959 TCTCTCAAAACAA.....GCCCTC 2978
55 ThrGluProLeuThrMetGlyLysMetSerAsn..... 65
2979 CAGGCCCTCTCAGACTCTCAAGAGGAAATGTCTCCCGCCGACCCG 3028
66 .....ValVal. 67
3029 TGTCTCTGTCCTCCAGAAAGCCAGCCAGCTGCACCTTAAGAGTCAAG 3078
68 LysLysTrpThrThrCysProSerAsnThrValLysThrGluAsnAlaThr 84
: |||||
3079 CCGGAGACAGACACCTGTGTGCGCAGAGCCCTCCCAAAAGCCAGCAGTAC 3128
84 r.....GlyProGluGluLeuGlyLeuProLeuGln..... 94
: |||
3129 CAAGAGCAACACCAAGAACTCTCCATCCCAAGAGGAGAGAGTAGAGG 3178
95 .....ArgSerLysSerGluHisLeuGlyTrp..... 103
3179 GGAAGGCTCCAGAGAGCTCTCGAGCAGACAGAGGTTCTCCGAGATACT 3228
104 .....PheProThrAspLeuPheAlaCysSerGluSerLeuArgAs 117
3229 GCAATCCTTTCCAGTG.....CCTCTTGGCAAA 3260
117 nGlyAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsn 134
3261 TGGTAAGCTCT.....AAACAGGAGAAAGCTCAAGTAGAATTGACA 3301
134 yLysIleSerLeuLeuHis.....SerSerLys 143
: |||||
3302 AACAAACAGACAGACCTTACATGAGGAGGAGCAAAAAGATGACGAGAA 3351
144 GluLysLeuArgGluArgLysLysLysLysLysLysLysLysLys 160
3352 GCAGAGTTAATGACGAGCAGGGGTGGAAAGGCTTTAATGACTGGAAC 3401
160 rLeuLeuProLysValLys.....GlyA 168
3402 CGCTGTGCTTCATATGAGTGGGAATTCGACAGAGTGTGAAGGCACT 3451
168 rLysAsnAspAlaAlaSerValLeuGluAlaThrValAspLysValLys 184
3452 CATCCAAAGTCACTTACTGTCTACTCAGAAAGTGTAGATCTCATTA 3501
185 TyrIle.....ArgG1 188
3502 TTGATTAATGTCATTAATTCCTTTCAGATGCCACAGGCCCAACACAA 3551

```

```

188 uLySleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSera 205
||||| : : : : : ||| : : :
3552 GAAATATTTCTCTTATGCAATGCGTGGCAGTCCATTGTAACATGG 3601
205 smMetArgPheCysLysLysGlnInThrProIleGluLeuSerLeuPro 221
||| : : : : : ||| : : :
3602 CGATGTTCTGTTAAAGAACAC.....ATAGCAATTAAGATATCT 3642
222 GLyThrValMetAlaGlnArgGluAsnSerValMetSerThrYrSerPr 238
||||| : : : : : ||| : : :
3643 CGTACTCTTAATAACACTTCGAGAGTCTTCCAAAGTCCGCGCAGGCAC 3692
238 oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSer 255
| : : : : : | : : : : :
3693 T.....TCTCCATGCATTGCAAGCACAGGCA 3718
255 hrProAspAlaGlnSerSerLeuAspGluAlaValArgValProSerSer 271
||||| : : : : : ||| : : :
3719 CACCATCCCTCTTCCCAATGCTTCTCTGCGCAGCTCCGAGGTC 3768
272 SerAlaSerGluAsnAlaIleGlyAspProTyrTrpThrHisLeuSer 288
||||| : : : : : ||| : : :
3769 CAGCAAGTCTGCGAGTGGGAGC.....AGTGG 3800
288 rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrTrpSerLys 305
||||| : : : : : ||| : : :
3801 GGTGGCTGCCATCTACAGACCCGACATCAG.....ATA 3841
305 aLhrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerL 321
||||| : : : : : ||| : : :
3842 TGACATCTCTCTAT.....GTACACATC 3864
322 .HisLeuProSerAlaMetProVal..... 330
||||| : : : : : ||| : : :
3865 ACAT..CCATGTTCTTACCGCTTTGACCTTGGACAGCCGAGGC 3911
331 .SerSerPheSerLeuGlyThrAlaLeuLeuGlyTyrPalaArgAla 346
||| : : : : : ||| : : :
3912 CCTCAGCAGAGAAATTAAGATCTTGTGCGGCTCAGCAATATGCT 3961
347 LeuHisIlePro.....ThrValCysAsnSerPheG 357
||||| : : : : : ||| : : :
3962 GCACCTTGCCCTCAACAGCAGTTGTGTGACCTGTGCACTATACAGA 4011
357 YArgIleLysSerThr 362
||||| : : : : : ||| : : :
4012 CAGGCTTTCAGCAGC 4027

seq_name: /cgn2_6/ptodata/1/lna/6_COMB.seq:us-08-545-860D-25
seq_documentation_block:
: Sequence 25, Application US/08545860D
: Patent No. 6040140
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: For Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545,860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE: 22-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/10930
: FILING DATE: 09-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/327,392
: FILING DATE: 19-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/320,559
: FILING DATE: 11-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/062,443
: FILING DATE: 14-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,094
: FILING DATE: 30-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/888,839
: FILING DATE: 27-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/805,093
: FILING DATE: 11-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TJU-1262
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9391 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 421..4053
: US-08-545-860D-25

alignment_scores:
: Quality: 113.00 Length: 424
: Ratio: 0.589 Gaps: 22
: Percent Similarity: 45.283 Percent Identity: 21.462

alignment_block:
US-09-389-000-2 x US-08-545-860D-25 ..
Align seg 1/1 to: US-08-545-860D-25 from: 1 to: 9391

25 LySLeYsAsnThrHisSer.LeuPheValPheIleIleProGluAsn 41
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2909 AGAAGCAATCAATCAACAGTCTTCACTTCTTCATCTCCCAAGAA 2958
41 he.....LySGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2959 TCTTCTAAACAAA.....GCCCTC 2978
55 ThrGluProLeuThrMetGlyLysMetSerAsn..... 65
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2979 CAGGCCCTCTCTACACTCCCAAGAAAGAAATGCTCCCGCCACCG 3028
66 .....ValVal. 67
3029 TGCTCTGCTCTCCAGAACGACCAAGCTTCGACTTAAGAGTCAAG 3078

```


68	LYSTIPRTPTHTHThCysProSerAsnThrValLysThrGluAsnAlaIth	84
3079	CGGGAAAGCAGACACCTGTGTGCCAGGACCTCCCAAAAGTCCAGCAGTAC	3128
84	r.....GlyProGluGluLeuGlyLeuProGluIn.....	94
3129	CAAGACACACCAAAAGACTCTTCCATTTCCCAAGCAGAGAGAGTAGAGG	3178
95ArgSerTyrSerGluHisLeuGlyTyr.....	103
3229	GCAAATCCTTTTCCAGTG.....CCTCTTGTCCAAA	3260
117	ngIysanGlyLeuGluLeuLeuAsnAlaSerLeuSerGluPheGluYAsnL	134
3261	TGGTAACTCT.....AAACCAGGCAAGCCTCAAGTGAAGTTTGACA	3301
134	ysLysIleSerLeuLeuHis.....SerSerLys	143
3302	AACAAACAAACAGACCTTCCATGAGGAGCGCAAAAAGATGAAAGCAAA	3351
144	GluYIysLeuArgArgGluArgIleLysTyrCysCysGluInLeuArgTh	160
3352	GCAGATTATATACGACGACAGGGTTGGAAAGGCTTTTAAGTACGTGAAC	3401
160	rLeuLeuProTyrValLys.....GlyA	168
3402	CGCTTGTCTCTTCACTGAGTGGGGAATGGCCAGAGTCTGAAACCACT	3451
168	rGlyAsnAspAlaIaSerValLeuGluAlaIthValAspTyrValYs	184
3452	CATCCAGACGACTTACTCTGTCTACTCGAAGACTGATAGTCTACTTAA	3501
185	TYRILE.....ArgI	188
3502	TTTCATATGTCAATTAAATCCTTCTCAGATGCCACAGCGCCAAACAAGA	3551
188	uLysIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerA	205
3552	GAATAATTTTGTCTGTTTATGCATGGGTGCCAGTCCATTTTGAACATGG	3601
205	smMetArpPheCysLysLysGlnGlnThrProIleGluLeuSerLeuPro	221
3602	CGATGTTTGGTTGTAATAAAGC.....ATAGCAATAAAGTATCT	3642
222	GlyIthValMetAlaGlnArgGluAsnSerValMetSerIthTyrSerPr	238
3643	CGTACTCTTAATTAACACTTCGAGAGATCTTCCAAAGTCCGCCAGGCAAC	3692
238	oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerT	255
3693	T.....TCTCATGCACTTGCAGACAGCAAGCA	3718
255	hrProAspAlaGluSerSerLeuAspGluAlaValArgValProSerSer	271
3719	CACACACCCCTCTTCCCAAAAGCCTTCCCGCCAGCTCCGTAAGGTCC	3768
272	serAlaSerGluAsnAlaIleGlyAspProTyrLysIthHisIleSerSe	288
3769	CATCACTAGCTGCGCAGTGTGGGAGC.....AGTGG	3800
288	rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysV	305
3801	GGTGGCTGCACCTATACGACACCCAGTCCATCCATCCAG.....AATA	3841
305	alrhrProSerTyrAspAlaIthrAlaValIthrAsnGlnAsnIleSerIle	321
3842	TGACATCTTCTAT.....GTCCACATC	3864

[illegible]

```

alignment_scores:      Length: 424
                        Quality: 113.00
                        Ratio: 0.889
                        Gaps: 22
Percent Similarity:    45.283
Percent Identity:      21.462

alignment_block:
US-09-383-000-2 x PCR-US94-04496-25 ..

Align seg 1/1 to: PCR-US94-04496-25 from: 1 to: 9391

```



```

REFERENCE/DOCKET NUMBER: KUBOTA-6A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1704
US-08-528-199-5

Alignment_scores:
Quality: 109.50      Length: 351
Ratio: 0.730         Gaps: 14
Percent Similarity: 42.735      Percent Identity: 20.513

alignment_block:
US-09-389-000-2 x US-08-528-199-5 ..
Align seg 1/1 to: US-08-528-199-5 from: 1 to: 1704

44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluPro..... 57
||| ||||| ||| :::: ||| |||
680 TGCTCAAGCGGCTGCGCGCTTCGTCGACGACAACTACCCGACGCGGTG 729
58 .LeuThrMetGluysMetSerAsnValValLysTyrTrpThrTrpCysP 74
::::: ||||| :::: ||||| :::: |||||
730 CTGCTGTCAGAGCGCAACAGTGGCCGACGACGCGGTGAGTACTTCGG 779
74 roSerAsnThrValLysThrGluAsnAlaThrGlyPro.GluGluLeuG1 90
::::: ||||| :::: ||||| :::: |||||
780 GCCGAGAGCGCTGACGACGCGCTGCGCGCCGACGAGTCAATCAGG 829
90 yLeuProLeuGlnArgSerTyrSerGlnHisLeuGlyTyrPhePro.Thr 106
||||| ||||| :::: ||||| :::: |||||
830 CCTTCCACTTCCCGGTGATGCCGCGCATCTTCATGCGCGTCCGCGCAG 879
107 AspLeuPheAlaCysSerGluSerLeuArgAsnGlyLeuGluLe 123
::::: ||||| :::: ||||| :::: |||||
880 TCGCGCTTCCGATCTCGAGATCATGAGACGACGCGCGGATCCCGGA 929
123 uAsnAlaSerLeuSerGluPheGluLysAsnLysLysIleSerLeuLeu 140
::::: ||||| :::: ||||| :::: |||||
930 GCGCTGCAGTGGGGCATCTTCTGCGCACAACGACGAGTCACTCAG 979
140 tSerSerLysGluLysLeuArgArgGluArgGlyLysTyrCysCysGlu 156
::::: ||||| :::: ||||| :::: |||||
980 AGATGGTCCACGACGAGGACCGC..... 1002
157 GlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAlaAl 173
1003 .....GACTACATGTGGGCG..... 1017
173 aserValLeuGlnAlaThrValAspTyrValLysTyrIleArgLysI 190
::::: ||||| :::: ||||| :::: |||||
1018 .....GAGTACGCCAAGACCCCGCATGAAG 1045
190 leSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMet 206
::::: ||||| :::: ||||| :::: |||||
1046 CCAACATCGGATCCGCGCGGCTCGCGCGCTGTCGACACGACGACG 1095
207 ArgPheCysLysLysGlnGlnThrProIleLeuLeuSerLeuProGly 223
::::: ||||| :::: ||||| :::: |||||
1096 AACCAAGATC...GAGCTGTTCACCGCGCTGCTGTCGTCGCGGCG 1140
223 rValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerProGlu 240

```

```

1141 ..... ||| :::: ||||| :::: |||||
240 rgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerThrPro 256
||| :::: ||||| :::: |||||
1169 TCGGC.....ATGGCGACACAATCTGGCTCGGTGACCGC..... 1203
257 AspAlaGluSerSerLeuAspGluAlaValArgValProSerSerAl 273
::::: ||||| :::: ||||| :::: |||||
1204 .....GACGGCTGGTACGCGCATCGAGCGAC 1232
273 aserGluAsnAlaIleGlyAspProTyrLysThrHisIleSerSerAla 290
::::: ||||| :::: ||||| :::: |||||
1233 CCGGACCGCAACTCGGC..... 1251
290 lalaSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThr 306
::::: ||||| :::: ||||| :::: |||||
1252 .....TTCGCGCGCCACG 1266
307 ProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisle 323
||| :::: ||||| :::: |||||
1267 CCGGCG.....AAGCTGCACCT 1283
323 uProSerAlaMetProProValSerSerPheSerLeu..... 335
||| :::: ||||| :::: |||||
1284 GCGAGCATCCAGGACCGGCTGACGGCTACACAGCGTCAAGTCGAGG 1333
336 .....GlyThrAlaLeuLeuGlyTyrTrpAlaArgAlaLeu 347
::::: ||||| :::: ||||| :::: |||||
1334 CGCAGCTGAGAACCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1383
348 HisIleProThrValCysAsnSerPheGlyArgIleLysSerThrCysle 364
||||| :::: ||||| :::: |||||
1384 CACATCCGCGCCGACGCGGACGCTTCGGC..... 1413
364 uLysPheThrLeuSerThrThrTyrTrpAlaGlnPheAspAsnLeuGly 380
||| :::: ||||| :::: |||||
1414 .....TTCGAGACACTCGC 1437

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-049-282B-1
seq_documentation_block:
; Sequence 1, Application US/08049282B
; Patent No. 5534660
; GENERAL INFORMATION:
; APPLICANT: CHUCK, George S.
; APPLICANT: DOONER, Hugo K.
; APPLICANT: COURTNEY-GUTTERSON, Neal
; APPLICANT: KELLER, Janis
; APPLICANT: NIJJAR, Charanjit S.
; APPLICANT: RALSTON, Edward J.
; TITLE OF INVENTION: PH GENES AND THEIR USES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourile and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,282B
; FILING DATE: 16-APR-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 12176-34

```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1149 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1146
US-08-049-282B-1

alignment_scores:
Quality: 108.50 Length: 306
Ratio: 0.714 Gaps: 15
Percent Similarity: 49.673 Percent Identity: 23.529

alignment_block:
US-09-389-000-2 x US-08-049-282B-1 ..
Align seg 1/1 to: US-08-049-282B-1 from: 1 to: 1149

6 ValProSerSerLeuSerAlaGluGluLeuGluAlaIleLeuLeuIleAr 22
   :::::::::::::: ::::::::::::::
187 TTACACGCGCGCCCTAGTATGATGATTA..... 216
   ::::::::::::::
22 gphegLyLysLysLysAsnThrHis...SerLeuPheValPheIleLeu 38
   :::::::::::::: ::::::::::::::
217 .....TCAAGAAGACACACATCTCTCAAAACAGTGCACCATTC 259
   ::::::::::::::
38 roGluAsnPhelys.....Gly 43
   ::::::::::::::
260 TTGACACCTCTCAACCAAGCTCCAAATTTCTTACCATATGCGC 309
   ::::::::::::::
44 CysIleSerGlnHisGlyMetAspIleAlaLeuThrGluProLeuThrMe 60
   ::::::::::::::
310 TGATATTC..... 318
   ::::::::::::::
60 tGluLysMetSerAsnValValLysTyrTrpThrCysProSerAsnT 77
   ::::::::::::::
319 .CAACAACCAATCTGCTTCACACAGTGGCCAGC...CCCACACCA 364
   ::::::::::::::
77 hrValLysThrGluAsnAlaThrGlyProGluGluLeuProLeu 93
   ::::::::::::::
365 CCGTCTCCAGCCCATTTCTGACGGCGCCACCTCCGGCCAGTGGCTG 414
   ::::::::::::::
94 GlnArgSerTyrSerGlnHisLeuGlyTyrPheProThrAspLeuPheAl 110
   ::::::::::::::
415 CTCGAAAGACATACATCTCTGTTCCATTTCTTACACTAAATACCAAC 464
   ::::::::::::::
110 acGysSerGln...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
   ::::::::::::::
465 TGCAGCTGAAGTTCTCCAAAGTCACGTGACGCTGACCACTGTGATTC 514
   ::::::::::::::
126 er...LeuSerGluPheGluLysAsnLysIle..... 136
   ::::::::::::::
515 CCACTGCATCTCGTTTCGAAAGGGTGTAGTATTAACACAAAGAACCT 564
   ::::::::::::::
137 SerLeuLeuHisSerSerLysGluLysLeuArgGluArgIleLysTyr 153
   ::::::::::::::
565 AGAGGAACATGATCTGCTGAACGACGCGCTGAGAAAAAGCTCAACA 614
   ::::::::::::::
153 rCysCysGluGlnLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
   ::::::::::::::
615 ACGATTATCATTTTGAAGTCACTGCTCTTTGTT.....ACGAAA 658
   ::::::::::::::
170 snAspAlaIleSerValLeuGluAlaThrValAspTyrValLysTyrIle 186
   ::::::::::::::
659 TGCATTAAGCTTCATCTTGTGTGACACCATATGTCACACAGCTTA 708
   ::::::::::::::
```

```
187 ArgGluLysIleSerPro...AlaValMetAlaGlnIleThrGluAlaIle 202
   ::::::::::::::
709 CGTAAAGAAAGTTCAGATCTTGAAGCTAGAGCCATCAGACGAGGCT... 756
   ::::::::::::::
202 uGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleGluLeu 219
   ::::::::::::::
757 .....ACCGTCACAGCAA 769
   ::::::::::::::
219 erLeuProGlyThrValMetAlaGlnArgGluAsnSerValMetSerThr 235
   ::::::::::::::
770 AGCATACAGGATGATGTAAGTGTTCACAGAAAGGCT..... 807
   ::::::::::::::
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsn 252
   ::::::::::::::
808 .....AAGAGAGATGAAGATGATGGAAGAAAGTGTGTGGAGG 848
   ::::::::::::::
252 Y.....CysSerThrProAspAlaGluSerSerLeuAspG 264
   ::::::::::::::
849 ACAGGCAAGATCAGCGCATCTTCGCCCTCACGACACATGA...GAGG 895
   ::::::::::::::
896 AGATAGTCAGATGAGATGATCAATTCAGAGATGATGATGATGAG 945
   ::::::::::::::
281 .....ProTyrLys 283
   ::::::::::::::
946 CTCAGGTGTCATACAAA 963
   ::::::::::::::

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-537-715-1
seq_documentation_block:
; Sequence 1, Application US/08537715
; Patent No. 5910627
; GENERAL INFORMATION:
; APPLICANT: Chuck, George S.
; APPLICANT: Dooner, Hugo K.
; APPLICANT: Courtney-Guterson, Neal
; APPLICANT: Keller, Janis
; APPLICANT: Nijjar, Charanjit S.
; APPLICANT: Ralston, Edward J.
; TITLE OF INVENTION: PH Genes and Their Uses
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,715
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,282
; FILING DATE: 16-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/04173
; FILING DATE: 15-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 012176-003410US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 1149 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1149
: OTHER INFORMATION: /note= "ph gene cDNA"
US-08-537-715-1

alignment_scores:
  quality: 108.50      length: 306
  ratio: 0.714        gaps: 15
  percent similarity: 49.673      percent identity: 23.529

alignment_block:
US-09-389-000-2 x US-08-537-715-1 ..

Align seg 1/1 to: US-08-537-715-1 from: 1 to: 1149

6 ValProSerSerLeuSerAlaGluGluLeuGluAlaIleLysLeuIleAr 22
  ::::::::::: ||| :::::::::::
187 TTACGAGGGGCGCTAGTATGATGATTA..... 216

22 GpHeLylsLysLysAsnThrHis...SerLeuPheValPheIleIleP 38
  ::::::::::: ||| |||
217 .....TCACAGAGAAGACACATTAATCTCAACAGTGCAGCATTC 259

38 roGluAsnPhelys.....Gly 43
  ||::::::::: ||
260 TTGAACACCTCTCAACCAAGCTCCAATTTCCTTCCATTAATGGC 309

44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluProLeuThr 60
  ||||||||
310 TGTATTCC..... 318

60 tGluLysMetSerAsnValIleLysTyrTrpThrCysProSerAsnT 77
  ::::::::::: ||| :::::::::::
319 .CAACCAACCCCAATCTGCTTCACAGGGGCCAGC...CCAGAGCA 364

77 hrValLysThrGluAsnAlaThrGlyProGluGluLeuGlyLeuPro 93
  ||||||||
365 CCGTCCAGCCCATTTCTTGACGGGGGCCACCTCCGGCCAGTGGCTG 414

94 GluArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuPhe 110
  ::::::::::: ||| :::::::::::
415 CTCAAAGACCATATCTCTGTCATTTCTTCACTAATAATACCAAC 464

110 acYserGlu...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
  ::::::::::: ||| :::::::::::
465 TGCAGCTGAAGTTTCTCCAAAGTCAGTGCCTACCACTGTTGATCT 514

126 er...LeuSerGluPheGlyLysAsnLysLysIle..... 136
  ||::::::::: ||| :::::::::::
515 CCACATGCATCTCGTTTGAAGAGGTGAGTAAACACAAAGAGCCT 564

137 SerLeuLeuHisSerSerLysGlyLysLeuArgGluArgIleLysTy 153
  ||| |||
565 AGTGAAACCATGCTCTGCTGACAGCGGCTAGAGAAACCTCAACGA 614

153 rCyserGluGluLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
  ||| |||
615 ACGGTTTTCATTTGAGGTCACTGTTCCTTTGTT.....ACGAAA 658

170 snAspAlaIleSerValLeuGluAlaThrValAspTyrValLysTyrI 186
  ||| |||
659 TGAATAAGCCTCATCTCTGTGTGACACCATATGATGCAAGCAGTTA 708

197 ArgGluLysIleSerPro...AlaValMetAlaGlnIleThrGluAla 202
  ||| |||

```

```

709 CETAAGAAAGTTACAGATCTTGAAGCTAGACCAATCAGAGGAGCT.. 756
202 uGlnSerAsnMetArgPheCysLysLysGlnIleThrProIleGluAs 219
757 .....ACGCTGCAGACAA 769

219 erLeuProGlyThrValMetAlaGlnArgGluAsnSerValMetSerThr 235
  ::::::::::: |||
770 AGCATACAGACTCTCTGAAAGCTGTGCAAGAGAGGCT..... 807

236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTyrPasn 252
  ::| | | | |
808 .....AAGAGGAGATGAGATAGTGAAGAGAGCTGTGTGTGAGAG 848

252 Y.....CysSerThrProAspAlaGluSerSerLeuAspG 264
  |||
849 ACAGGCAAGATCAGCGCATCTTCGCTCAACGACACATGAA...GAGG 895

264 LuAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGlyAs 280
  |||
896 AGATAGTCGAAGTGAAGATATCAATATCAGAGATGATGCTGTGAGAG 945

281 .....ProTyrLys 283
  |||
946 CTCAGGTGTCATACAAA 963

seq_name: /cgn2_6/ptoddata/1/ina/PCTUS_COMB.seq:PCT-US94-04173-1

seq_documentation_block:
: Sequence 1, Application PC/RUS9404173
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: PH GENES AND THEIR USES
: NUMBER OF SEQUENCES: 4
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04173
: FILING DATE: 15-APR-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 12176-34-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1149 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1146
: PCT-US94-04173-1

alignment_scores:
  quality: 108.50      length: 306
  ratio: 0.714        gaps: 15
  percent similarity: 49.673      percent identity: 23.529

alignment_block:
US-09-389-000-2 x PCT-US94-04173-1 ..

```

Align seg 1/1 to: PCT-US94-04173-1 from: 1 to: 1149

```

6 ValProSerSerLeuSerAlaGluGluLeuGluAlaAlaLeuLeuIleLeu 22
   ::::::::::: ||| :::::::::::
187 TTACGAGCGCGCCCTGATTGATGATTA..... 216
22 pPheGlyLysLysLysAsnThrHis...SerLeuPheValPheIleLeu 38
   ::::::::::: ||| :::::::::::
217 .....TCACAGAGACACACATTACTCTCAACAGAGTGCACATTC 259
38 roGluAsnPheLys.....Gly 43
   ||| :::::::::::
260 TTGAACACCTCTCAACACCAACCTCAATTTCTCTACCAATATGGCC 309
44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluProLeuThrMe 60
   ||| :::::::::::
310 TGTATTTCC..... 318
60 tGluLysMetSerAsnValAlaLysTyrTrpThrCysProSerAsn 77
   ::::::::::: ||| :::::::::::
319 .CAACAAACCAATCTGCCTTCAAGAGTGCCAGC...CCAGACACA 364
77 hValLysThrGluAsnAlaThrGlyProGluGluLeuProLeu 93
   ||| :::::::::::
365 CCGTCTCCAGCCCATTTCTTGACGGCGGCCACCTCCGCGCAGTGGCTG 414
94 GluArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuPheAl 110
   ::::::::::: ||| :::::::::::
415 CTCAAAACGATCTATCTCTCTCTTCATTTCTTCACACTCAATATACCAAC 464
110 aCysSerGlu...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
   ::::::::::: ||| :::::::::::
465 TGCACCTGAAGTTCTCCAAAGTCAAGTCAAGCTCACTGATGATCT 514
126 er...LeuSerGluPheGlyLysAsnLysLysIle..... 136
   ||| ::::::::::: ||| :::::::::::
515 CCACGTGATCTCTGCTTTCGAAAAGGCTGATGATTAACACAGAGAGCCT 564
137 SerLeuLeuHisSerSerLysGluLysLeuArgArgGluArgIleLys 153
   ||| ::::::::::: ||| :::::::::::
565 AGTGAACCAATGATCTGCTGACGACGCGCTAGAGAAAGCTCAACA 614
153 rCysCysGluGluLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
   ||| ::::::::::: ||| :::::::::::
615 ACGGTTATCATTTGAGGTGACTGCTGCTTTGTT.....ACGAAA 658
170 snAspAlaAlaSerValLeuGluAlaThrValAspTyrValLysTyrIle 186
   ||| ::::::::::: ||| :::::::::::
659 TGGATTAAGCCCTCATTTCTTGTCACACCATGAAATATGTCACAGCTTA 708
187 ArgGlyLysIleSerPro...AlaValMetAlaGluIleThrGluAlaLe 202
   ||| ::::::::::: ||| :::::::::::
709 CGTAAGAAAGTTCAGAGATCTTGAAGCTAGAGCCATCAGACGAGGCT.. 756
202 uGlnSerAsnMetArgPheCysLysLysGluGlnThrProIleGluLeu 219
   ::::::::::: ||| :::::::::::
757 .....ACGCTGCAGACAA 769
219 erLeuProGlyThrValMetAlaGluArgLysSerValMetSerThr 235
   ::::::::::: ||| :::::::::::
770 AGGATACAGAGTACTGTGAAGGTGTCAGAGAGAGGGT..... 807
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsn 252
   ::::::::::: ||| :::::::::::
808 .....AAGAGAGAAATGAAGATAGTGAAGAGAGTGTGGTGAAG 848
252 Y.....CysSerThrProAspAlaGluSerSerLeuAspG 264
   ::::::::::: ||| :::::::::::
849 ACAGGCAAGATCAGCGCATCTTCGCTCAACGACACATGAA...GAGG 895
264 lAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGlyAsp 280
   ||| ::::::::::: ||| :::::::::::

```

896 AGATAGTCAGAGTAGAAGATATCATATGACAGAGTATGATGATGAG 945

281ProTyrLys 283

946 CTCAGGTCTCATACAA 963

seq_name: /cgn2_6/prodata/1/lna/5b_COMB.seq:US-08-537-715-3

seq_documentation_block:

; Sequence 3, Application US/08537715

; Patent No. 5910627

; GENERAL INFORMATION:

; APPLICANT: Chuck, George S.

; APPLICANT: Dooner, Hugo K.

; APPLICANT: Courtney-Guterson, Neal

; APPLICANT: Keller, Janis

; APPLICANT: Nijjar, Charanjit S.

; TITLE OF INVENTION: PH Genes and Their Uses

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/537,715

; FILING DATE: 16-FEB-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/049,282

; FILING DATE: 16-APR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US94/04173

; FILING DATE: 15-APR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 012176-003410US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1857 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 481..1632

; OTHER INFORMATION: /note= "ph6 gene cDNA"

; US-08-537-715-3

alignment_scores:

Quality: 108.50

Ratio: 0.714

Percent Similarity: 49.673

Percent Identity: 23.529

alignment_block:

US-09-389-000-2 x US-08-537-715-3

Align seg 1/1 to: US-08-537-715-3 from: 1 to: 1857


```

77 hVallThrGluSnaIaThrGlyProGluLeuGlyLeuProLeu 93
   |||:::
848 CGGTCCAGCCCATTTCTGAGCGGGGCCACCTCGCGGCGTGCTG 897
94 GlnArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuHea 110
   ::::
898 CTCAAAGACATATCTCTGTCATTTCTTCACATMAATACCAAC 947
110 acysSerGlu...SerLeuArgasnGlyasnGlyLeuGluLeuSnaIas 126
   ::::
948 TGCAGCTGAAGTTCTCCAAAGCAGCTGACGCTACCACGTGATTCCT 997
126 er...LeuSerGluPheGluLysAsnLysLysLe... 136
   |||:::
998 CCAGTCATCTCGCTTCGAAAGGGGTAGTATTAACACAAAGAGCCT 1047
137 SerLeuLeuHisSerSerLysGlyLysLeuArgGluArgLysTyr 153
   |||:::
1048 AGTGAACCATGCTACTGCTGAAACGCGGTAGAGAAAGCTCAACGA 1097
153 rCysCysGluGlnLeuArgThrLeuLeuProTyrValLysGlyArgLysA 170
   |||:::
1098 AGCGTTATCATTTTGAAGTCACTTCTCTTGT...ACGAAA 1141
170 snAspAlaIaSerValLeuGluAlaThrValAspTyrValLysTyrIle 186
   |||:::
1142 TCGATMAACCTCCATCTTGTGTGACACCATAGATATGTCAAGCAGTTA 1191
187 ArgGlyLysIleSerPro...AlaValMetAlaGlnIleThrGluLala 202
   |||:::
1192 CATAAGAAAGTTGAGATCTTGAGCTAGACCAATCAGACGAGGCT.. 1239
202 ucInSerAsnMetArgPheCysLysGlnGlnInThrProIleGluLeuS 219
1240 .....ACGCTGCACAGCAA 1252
219 erLeuProGlyThrValMetAlaGlnArgLysSerValMetSerThr 235
   ::::
1253 ACGATACAGCTACTGTGAAGGTGTGCAAGAGAGGGT..... 1290
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrrpAsn 252
   ::::
1291 .....AAGAGGAAATGAGATGAGAGGAAGTGTGTGGTAGAG 1331
252 Y.....CysSerThrProAspAlaGlnSerSerLeuAspG 264
1332 ACAGGCAAGATCAGCGCATCTTCGCCCTCAAGACACATGAA...GAGG 1378
264 lAlaValArgValProSerSerSerAlaSerGluSnaIaIleGlyAsp 280
   |||:::
1379 ACATAGTCAGAGTAGAATATTCAGAGAGTATGATGATGATGAGAG 1428
281 .....ProTyrLys 283
1429 CTCAGGCTGCCATACAA 1446

seq_name: /cgn2_6/ptodate/1/ina/5A_COMB.seq:us-07-903-710-2
seq_documentation_block:
: Sequence 2, Application US/07903710
: Patent No. 5302519
: GENERAL INFORMATION:
: APPLICANT: Eyer,D.E.;Eisenman,R.N.;Blackwood,E.M.;Ayer,D.M.
: TITLE OF INVENTION: MAX: A HELIX-LOOP-HELIX ZIPPER PROTEIN THAT FORMS A
: TITLE OF INVENTION: SEQUENCE-SPECIFIC DNA-BINDING COMPLEX WITH MYC AND MAD
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Christensen, O'Connor, Johnson and Kindness
: STREET: 2800 Pacific First Center, 1420 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98101-2347

```

```

COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
: COMPUTER: IBM PC/386 Compatible
: OPERATING SYSTEM: MS-DOS 4.01
: SOFTWARE: Word for Windows-t
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/903,710
: FILING DATE: 19920623
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/756,195
: FILING DATE: 09/09/91
: ATTORNEY/AGENT INFORMATION:
: NAME: Sundsmo, John, S.
: REGISTRATION NUMBER: 34,446
: REFERENCE/DOCKET NUMBER: FHCR-1-6550
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 1-206-682-8100; 1-206-224-0727 (direct)
: TELEFAX: 1-206-224-0779
: TELEX: 4938023
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 510 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: DESCRIPTION: Human helix-loop-helix zipper protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens; Eukaryota; Animalia; Metazoa;
: ORGANISM: Chordata; Vertebrata; Mammalia; Theria; Eutheria;
: ORGANISM: Primates; Haplorhini; Catarrhini; Hominoidea;
: IMMEDIATE SOURCE: Human lymphoid B cell Manca cell line,
: US-07-903-710-2

alignment_scores:
: Quality: 108.00 Length: 131
: Ratio: 1.301 Gaps: 3
: Percent Similarity: 63.359 Percent Identity: 24.427

alignment_block:
: US-09-389-000-2 x US-07-903-710-2 ..
Align seg 1/1 to: US-07-903-710-2 from: 1 to: 510

140 HisSerSerLysGluLysLeuArgGluArgGlyIleLysTyrCysCysG 156
   |||:::
82 CATAAGCACTGGAAGCAAAAGGTAGGACACATCAAGACAGCTTTCA 131
156 ucInLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAla 173
132 CAGTTGGGAGACTCAGTCCATCTCAAGAGAGAGGATCCCGG 181
173 lAserValLeuGluAlaThrValAspTyrValLysTyrIleArgGlyLys 189
   |||:::
182 CCCAAATCTAGACAAAGCCACAGACTATATCCAGATATGCGAAGAAA 231
190 lIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMe 206
232 .....AACCAACACACACACAGCAAGATYT 254
206 tArgPheCysLysGlnGlnInThrProIleGluLeuSerLeuProGlyT 223
   |||:::
255 TGCAGCACTCAAGCGGACGAATGCTTCTGAGACGACAGTCCGCGAC 304
223 hValMetAlaGlnArgLysSerSerValMetSerThrTyrSerProGlu 239
   |||:::
305 TGGAGAAAGCG...AGGTCAAGTGCACCACTGACAGCAACTACCCCTCC 351
240 ArgGlyLeuGlnPheLeuThrAsn.....ThCysTrr 250
   ::::

```


THIS PAGE BLANK (USPTO)

heat shock protein 86 - mouse
N:Alternate names: HSP86
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 22-Jun-1999
C:Accession: B32848; PS0021; A40338; A37345
R:Moore, S.K.; Kozak, C.; Robinson, E.A.; Ullrich, S.J.; Appella, E.
J. Biol. Chem. 264, 5343-5351, 1989
A>Title: Murine 86- and 84-kDa heat shock proteins, cDNA sequences, chromosome assignment
A:Reference number: A32848; MUID:89174568
A:Accession: B32848
A:Molecule type: mRNA
A:Residues: 1-733 <MR1>
A:Cross-references: GB:04633; NID:g194030; PIDN:AAA53068.1; PID:g309318
R:Hoffmann, T.; Hovemann, B.
Gene 74, 491-501, 1988
A>Title: Heat-shock proteins, Hsp84 and Hsp86, of mice and men: two related genes encoded
A:Reference number: PS0020; MUID:89232740
A:Accession: PS0021
A:Molecule type: mRNA
A:Residues: 6,'A','8'-242,247-355,'K' <HOE>
A:Cross-references: GB:MX3630; NID:g194032; PIDN:AA37868.1; PID:g194033
R:Moore, S.K.; Appella, E.; Villar, C.J.; Kozak, C.A.
Genomics 10, 1019-1029, 1991
A>Title: Mapping of the mouse 86-kDa heat-shock protein expressed gene (Hsp86-1) on chro
A:Reference number: A40338; MUID:92009901
A:Accession: A40338
A:Molecule type: DNA
A:Residues: 556-634 <MR2>
A:Cross-references: GB:M57673; NID:g194028; PIDN:AA37867.1; PID:g194029
R:Legoux, V.; Megier, V.; Ouelard, C.; Barnier, J.V.; Bensade, O.; Morange, M.
Differentiation 41, 42-48, 1989
A>Title: High constitutive transcription of HSP86 gene in murine embryonal carcinoma cell
A:Reference number: A37345; MUID:90033873
A:Accession: A37345
A:Molecule type: mRNA
A:Residues: 460-733 <LEB>
A:Cross-references: GB:X16857; NID:g51456; PIDN:CAA34748.1; PID:g51457
C:Comment: In response to temperature stress, to treatment with certain chemicals and am
molecular-weight proteins.
C:Comment: This protein is one of two forms of 80-90 kDa heat shock proteins found in m
C:Genetics:
A:Gene: HSP86
A:Introns: 586/3
A:Notes: the list of introns may be incomplete
C:Superfamily: heat shock protein 90
C:Keywords: estrogen-induced protein; heat shock; phosphoprotein; steroid receptor comp
F:223-323/Region: heat shock protein 86 #status experimental <AT>
F:523-570/Region: highly charged

```

Query Match          5.5%  Score 115.5:  DB 1;  Length 733;
Best Local Similarity 22.8%  Pred. No. 0.99;
Matches 87;  Conservative 57;  Mismatches 126;  Indels 111;  Gaps 22;

QY      6  VPSLSAELELEIKIRRGKK---KNTHSLFVFI--PENFKGCSIGHOMDIAL---TEP  57
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      395  LPLNISRRLQOSKILKXIRKLVKCCLEFLFELHEDKDKENKFFEQSKNKLGHEDS  454

QY      58  LTMERMSNVYKIWTQCPSNVTYKTENATGPEELGL---PLQRSYSEHLGYEPT---DLF  109
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      455  QNRKRLSELRLRYT-----SASGDGEMSLDYDCYRMRKNQKH--YFYTGTQDOV  503

QY      110 ACS---ESLRNNGLEL-----NASISEPEKKKKIS-----LHSSKEKUR  148
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      504  ANSAFVERLR--HGLEVIYIMEPIDEYCVQOQKEFEKGLTVSVTKEGELPEDEEKKQ  562

QY      149  ERIKCCCEQLRTLLPYVYGRKNDASVLEAVDYVKYIREKI--SPAVNAQITREALQSNR  207
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      563  EKKKKRPFENLCKIM-----KDILEKKYKKVYVSNRLVTSPPCCIVTSTYGTANME  612

QY      208  FCKKOQPTIEUSLPETVMAQRENSVMSTVSPERGLOFLNTTCWNGCSPPDAESSIDENVR  267
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

```

```

Db 613 RIMRAQ-----ALPDNSTMGYMAKKHLEI-----NPD-HSIIETLR 648
QY 268 VPS-SSASENAIGD---PYKTHISSAALSIN-SLHTVRYSKV-----TFSYDA 311
Db 649 OKAEADNRKDKSKVLVILLITALLSSGFSLEDPQTHANRYIRKIKGLGIDEDDPYVD 708
QY 312 T-AVTNQNISIHLPSPMPV 350
Db 709 TSAAVTEER-----MPL 720

```

RESULT 3
T31095
vitellogenin precursor - *Oreochromis aureus*
C:Species: *Oreochromis aureus*
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: T31095
R:LilM, E.H.; Lam, T.J.; Ding, J.L.
submitted to the EMBL Data Library, August 1997
A:Description: Cloning of full-length *Oreochromis aureus* vitellogenin cDNA and its de
A:Reference number: Z20978
A:Accession: T31095
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1788 <LIM>
A:Cross-references: EMBL:AF017250; NID:g4102880; PID:g4102881; PIDN:ABD01515.1
A:Superfamily: vitellogenin

```

Query Match          5 4%: Score 114; DB 2; Length 1788;
Best Local Similarity 20.9%; Pred. No. 4, 6;
Matches 80; Conservative 53; Mismatches 140; Indels 110; Gaps 19;

QY 5 KVPSSLSAEELBAIKLI-----RFGKKNTSLFVFIIPENFGKCSIGHGMIDALT 55
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1418 KLPGSIKHAKRALKSLIPIAEQGVNRYAKKN-----PRN-----QIKLT 1458
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 56 ELLTMEKSNVYKWTTCGTSNVTKTEMATGPELGLPL-----ORSSEHLGYFP 105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1459 VAVATETSNNI---LNPKAIVYKRGVCLP--VALPIGNTAAELQATRDNNADKKSYLV 1513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 TDLFACSESLRNGNGLLENLASLSEFEKN-KISLHSSREKTLRERIKYCCEDLRLLPY 164
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1514 TKANAVECSL-----INNTLTFFNNRKARDELPHSCYOVLAD-----CTPELKFVLL 1562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 VKGRND--ASVLEATVYVYKIREKISPAVMAQTTELQASMRCKKQQTIELSLPS 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1563 KKDQIQDONQILVKTISDDVDVI--RNNNAIAVAVNGVELIPNSLPI-----LHPSG 1612
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 TVMAORENSVMSVTSAPERLQ--FL-----TNTCWNGCSPTPDASSIDEAV 266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1613 NIHIQSNBEGITLNPAPSHQLGEVFLGFNLERYKAVADMKGKTC-GACGT--ASGVNGDEX 1669
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 RVPSSSAEENALG-----DPKTHISSAALPL-----NSLHTRYYS 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1670 KTPSQVTKDAISYHSHWVLSNTRCDPSECSIKQESVLEKRVIEGVESKCYSEVPVL 1729
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1730 QCLPCCIPVRTTVVAVGFHCLPS 1752
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
HHH086
heat shock protein 90-alpha - human
N:Alternate names: heat shock protein 86; heat shock protein 89 alpha
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1990 #sequence revision 30-Sep-1990 #text change 22-Jun-1999
C:Accession: A32319; J00128; PS0020; S06988; J00724; B31420; S67961
R:Hitkey: E.; Brandon, S.E.; Smale, G.; Lloyd, D.; Weber, L.A.
Mol. Cell. Biol. 9, 2615-2626, 1989
A:Title: Sequence and regulation of a gene encoding a human 89-kilodalton heat shock
I:Reference number: A32319; MUID:89343979

```

A:Accession: A32319
 A:Molecule type: DNA
 A:Residues: 1-732 <HIC>
 A:Cross-references: GB:M20724; NID:9341598; PIDN:AAA63194.1; PID:9703087
 A>Note: the authors translated the codon AAC for residue 383 as Asp
 R:Walter, T.; Drabant, B.; Krebs, H.; Tomalak, M.; Heiss, S.; Benecke, B.J.
 Gene 83, 103-115, 1989
 A:Title: Cloning and analysis of a human 86-kDa heat-shock-protein-encoding gene.
 A:Reference number: J00128; MUID:90076956
 A:Accession: J00128
 A:Molecule type: DNA
 A:Residues: 1-312 <MAL>
 A:Cross-references: GB:M30626; NID:9184418; PIDN:AAA36023.1; PID:9184419
 A>Note: the authors translated the codon AAA for residue 58 as Leu and AAC for residue 8
 R:Hoffmann, T.; Hovemann, B.
 Gene 74, 491-501, 1988
 A:Title: Heat-shock proteins, Hsp84 and Hsp86, of mice and men: two related genes encode
 A:Reference number: P50020; MUID:89232740
 A:Accession: P50020
 A:Molecule type: mRNA
 A:Residues: 1-312 <HOF>
 A:Cross-references: GB:X07270; NID:932485; PIDN:CAA30255.1; PID:932486
 R:Yamazaki, M.; Akao, K.; Miwa, T.; Imai, T.; Soeda, E.; Yokoyama, K.
 Nucleic Acids Res. 17, 7108, 1989
 A:Title: Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock protein from hu
 A:Reference number: S06898; MUID:89386066
 A:Accession: S06898
 A:Molecule type: mRNA
 A:Residues: 1-20, 'LSG', 25-62, 'T', 64-161, 'GVLSR', 168-732 <YAM>
 A:Experimental source: peripheral blood lymphocytes
 R:Yamazaki, M.; Tashiro, H.; Yokoyama, K.; Soeda, E.
 Agric. Biol. Chem. 54, 3163-3170, 1990
 A:Title: Molecular cloning of cDNA encoding a human heat-shock protein whose expression
 A:Reference number: J00724; MUID:91242090
 A:Accession: J00724
 A:Molecule type: mRNA
 A:Residues: 1-62, 'T', 64-732 <YAZ>
 A:Experimental source: adenovirus type 12 E1A-transfected HeLa cells
 R:Lees-Miller, S.P.; Anderson, C.W.
 J. Biol. Chem. 264, 2431-2437, 1989
 A:Title: Two human 90-kDa heat shock proteins are phosphorylated in vivo at conserved se
 A:Reference number: A92741; MUID:89123325
 A:Accession: B31420
 A:Molecule type: protein
 A:Residues: 2-21, 225-274 <LEB>
 A:Experimental source: HeLa cells in exponential growth
 A>Note: phosphorylation sites were determined by metabolic labelling with [32-P]orthopho
 R:Nemoto, T.; Ohara-Nemoto, Y.; Ota, M.; Takagi, T.; Yokoyama, K.
 Eur. J. Biochem. 233, 1-8, 1995
 A:Title: Mechanism of dimer formation of the 90-kDa heat-shock protein.
 A:Reference number: S67961; MUID:96061925
 A:Accession: S67961
 A:Molecule type: protein
 A:Residues: 533-539, 542-551, 'AO', 619-634 <NEM>
 C:Comment: in response to temperature stress, to treatment with certain chemicals and am
 molecular-weight proteins.
 C:Genetics:
 A:Gene: GDB:HSPCA; HSPC1
 A:Cross-references: GDB:118813; OMIM:140571
 A:Map position: Xpter-Xq22
 A:Introns: 54/3; 177/1; 221/3; 327/3; 383/1; 446/3; 496/1; 585/3; 697/1
 C:Superfamily: heat shock protein 90
 C:Keywords: estrogen-induced protein; heat shock; phosphoprotein; steroid receptor comp
 F:223-322/Region: highly charged
 F:534-569/Region: highly charged
 F:231,263/Binding site: phosphate (Ser) (covalent) #status experimental.

Query Match 5.4%; Score 113.5; DB 1; Length 732;
 Best Local Similarity 21.8%; Pred. No. 1.4;
 Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;

6 VPSLSAELEAIKIRFGK--KNTHSLFVIT--PENFKCISGKMDIAL---TEP 57

DB 394 LPLNLSREMLQOSKILKYLKRLVKKLELFTLEADKENVKPFEOSSKNKLGIHEDS 453
 58 LMEKMSNVKWKTCPSNTVTENATGPEELGL----PLQSYSEHLGYPTDL----- 108
 DB 454 QNRKILSELRLRYT-----SASGDEMVLNDYCRKMKENKHIIYITGFKRDVA 503
 QY 109 -FACSESLRNGLEL-----NASLSEFEKNKIS-----LHSSKEKLRE 149
 DB 504 NSAFVEYRLRK-HGLEVIYIWEPIDEYCVQQLKEFGKTLVSTKGLPEDEEEKKOE 562
 QY 150 RIKYCCOELRTLLPYVKGKNDAAVLEATVDYVYIRKTI-SPVAMQITLALSNMNF 208
 DB 563 EKRTFENLCKIM-----KDLKKYKVVVSNELVSPCCIVTSTGTWAMNER 612
 QY 209 CKKQOPIELSLPGVMAQRENSVMSTVSPEERLOFLTNCNGCSTPDASSLDEAVR 268
 DB 613 IMKAQ-----ALRDNSTMGYMAKKHLEL-----NPD--HSIETLRQ 648
 QY 269 PS-SSASENAIGD---PYKTHISSAALSLN--SLHVRVYS--KVTPSYDATTATNOMI 319
 DB 649 KAEADKNDKSVKDLVILLYETALLSGFLEDPQTHANRYMKILGIGIDEDPTADPT 708
 QY 320 SIHLDSAMPV 330
 DB 709 SAAYTEEMPL 719

RESULT 5
 747987
 hypothetical protein F21F14.120 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 28-Jul-2000
 C:Accession: T47987
 R:Choisne, N.; Robert, C.; Brotlier, P.; Winkler, P.; Cattolico, L.; Artiguenave, F.;
 submitted to the Protein Sequence Database, February 2000
 A:Reference number: Z24461
 A:Accession: T47987
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <CHO>
 A:Cross-references: EMBL:AL138642
 A:Experimental source: cultivar Columbia; BAC clone F21F14
 C:Genetics:
 A:Map position: 3
 A:Introns: 11/3; 209/3; 332/3
 A>Note: F21F14.120
 C:Superfamily: Arabidopsis thaliana hypothetical protein F21F14.120

Query Match 5.3%; Score 112; DB 2; Length 359;
 Best Local Similarity 22.3%; Pred. No. 0.65;
 Matches 65; Conservative 43; Mismatches 107; Indels 76; Gaps 12;

QY 82 NATGPEELGLP-----QRSYSEHLGYPTDLFACSESLRNGLELNASLSEFEKNKI 136
 DB 125 NMTLPSTSSPLSARRRKRKINHL--LPQENTREKRRKRKPKKNMEI---ENQRI 178
 QY 137 SLHSSGKELRREIRIKYCCOELRTLLPYVKGKNDAAVLEATVDYVYIRKISPAVVA 196
 DB 179 N-HIAVERNRRKQMEHNSLRALLPSPYIORQOASIVGAIINYK-----VLE 227
 QY 197 QTTEALQSNMRECKKQOPIELSLPGVMAQRENSVMSTVSPEERLOFLTNCNGCSTP 256
 DB 228 QTIQSLHSQKR-----TQOQNSSEV---VENALNHU-----SGISSN 261
 QY 257 DAESSLDEAVRPSSAS-----ENAGIDPKYTHISSAALSLNSIH---TYRY 301
 DB 262 DMTLTLEDQTCIPKIEATYIQNHVSLKYQCEKKQQLKGLISLEKRLTYVAHLNITSS 321
 QY 302 YSKVTPSYDATTATNOMISIHPSAMPVPSFSLSGTALGNARALHPTV 352
 DB 322 HSSVSTSFNWKQMEDE---CDLESA-----DEITAAVHRIFDIPPTI 359

Db 559 EKKAKFENLCKIN-----KDILEKKREKVVSNRLVTSPPCITVSTYWTANMER 608
 QY 209 CKKOQPIELSLPGTVMAORENSVMSTYSPERGLOFLTNCWNGCSTPDASSLDAVRV 268
 Db 609 IMKAO-----ALRDNSTMGYMAKKLEI-----NPD--HSIIETLRQ 644
 QY 269 PS-SSASENAIGD---PYKTHISSAALSIN--SLHTVRYYS--KYTPSYDATAVNTQNT 319
 Db 645 KAEADKNDKSVKNDLIVLTETALLSSGFSLEDPQTHANRITRYIKIKGLGIDEDDTAAEEA 704
 QY 320 SIHLPSAMPV 330
 Db 705 SPAYTEMPPL 715

RESULT 9
 B38431
 Myc-binding factor Max, short form - human
 C:Species: Homo sapiens (man)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Nov-1999
 C:Accession: B38431
 R:Blackwood, E.M.; Eisenman, R.N.
 A:Title: Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding
 A:Reference number: A38431; MUID:91173288
 A:Accession: B38431
 A:Molecule type: mRNA
 A:Residues: 1-151 <BLA>
 A:Cross-references: GB:M64240; NID:q187390; PIDN:AAA6200.1; PID:q187391
 A:Note: the authors translated the codon GAG for residue 94 as Gly

Query Match 5.2%; Score 108; DB 2; Length 151;
 Best Local Similarity 24.4%; Pred. No. 0.37; Length 160;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKERLRERIRKYCCQDLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVNAQIT 199
 Db 19 HNALEKRRDHIDKDSFSLSDVPSLOGEKASRAQILDKATEYIQYMRK-----N 69
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 70 HTHQODIDDLKRONALLEQOVRALEKA-RSSAQLQTNYPSSDNLXTNAGSTISAFDGG 128
 QY 254 STPDASSLDE 264
 Db 129 SDSSSESEPEE 139

RESULT 10
 I50379
 gene max protein - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: I50379
 R:Solleberger, K.G.; Kao, T.L.; Taparowsky, E.J.
 O:Oncogene 9, 661-664, 1994
 A:Title: Structural analysis of the chicken max gene.
 A:Reference number: I50379; MUID:94119609
 A:Accession: I50379
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-160 <SOL>
 A:Cross-references: GB:U12469; NID:q414723; PIDN:AAA16834.1; PID:q414724
 C:Genetics:
 A:Gene: max
 A:Introns: 12/3; 21/3; 57/3; 99/1

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 23.7%; Pred. No. 0.4;
 Matches 31; Conservative 26; Mismatches 58; Indels 16; Gaps 3;

QY 140 HSKERLRERIRKYCCQDLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKDSFSLSDVPSLOGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQOVRALEKARSSAQLQANV-PAADSSLYTNKGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

RESULT 11
 S33118
 max protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S33118
 R:Vaestrik, I.; Koskinen, P.J.; Altalo, R.; Maekelae, T.P.
 O:Oncogene 8, 503-507, 1993
 A:Title: Alternative mRNA forms and open reading frames of the max gene.
 A:Reference number: S33118; MUID:93149613
 A:Accession: S33118
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <VAE>
 A:Cross-references: EMBL:X66867; NID:g34469; PIDN:CAA47337.1; PID:g34470
 C:Genetics:
 A:Keywords: DNA binding; transcription regulation

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.4; Length 160;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKERLRERIRKYCCQDLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKDSFSLSDVPSLOGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQOVRALEKA-RSSAQLQTNYPSSDNLXTNAGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

RESULT 12
 A38431
 Myc-binding factor Max, long form - human
 C:Species: Homo sapiens (man)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 23-Mar-1993
 C:Accession: A38431
 R:Blackwood, E.M.; Eisenman, R.N.
 O:Science 251, 1211-1216, 1991
 A:Title: Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding
 A:Reference number: A38431; MUID:91173288
 A:Accession: A38431
 A:Molecule type: mRNA
 A:Residues: 1-160 <BLA>
 A:Cross-references: GB:M64240
 A:Note: the authors translated the codon GAG for residue 103 as Gly

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.4;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKERLRERIRKYCCQDLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKDSFSLSDVPSLOGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQOVRALEKA-RSSAQLQTNYPSSDNLXTNAGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

Db 28 HNALEKRRDHKIDKSFHSLRDSVPSLOGEKASRAQILDKATEYIOYMRK-----N 78

QY 200 EALQSNMRFCKKQOTPELSLPGTVMAORENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLLEQVRALEKA-RSSAQLQTNPPSSDNLTYNAKGTISAFDGG 137

QY 254 STPDSESLDE 264

Db 138 SDSSESEPEE 148

RESULT 13

A38488

Myn protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 31-Jul-1991 #sequence_revision 31-Jul-1991 #text_change 16-Feb-1997

C:Accession: A38488

R:Prendergast, G.C.; Lave, D.; Ziff, E.B.

Cell 65, 395-407, 1991

A:Title: Association of Myn, the murine homolog of max, with c-Myc stimulates methylatid

A:Reference number: A38488; MUID:91208676

A:Accession: A38488

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-160 <PRE>

A:Cross-references: GB:M63903

C:Keywords: DNA binding; transcription regulation

Query Match 5.2%; Score 108; DB 2; Length 160;

Best Local Similarity 24.4%; Pred. No. 0.4;

Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKSEKLRERIKYCCCEQLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVMAQIT 199

Db 28 HNALEKRRDHKIDKSFHSLRDSVPSLOGEKASRAQILDKATEYIOYMRK-----N 78

QY 200 EALQSNMRFCKKQOTPELSLPGTVMAORENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLLEQVRALEKA-RSSAQLQTNPPSSDNLTYNAKGTISAFDGG 137

QY 254 STPDSESLDE 264

Db 138 SDSSESEPEE 148

RESULT 14

S39792

Protein max - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995

C:Accession: S39792

R:Itawa, M.

Biochim. Biophys. Acta 1216, 492-494, 1993

A:Title: Molecular cloning and sequencing of rat Max cDNA: castration-induced expression

A:Reference number: S39792; MUID:94092746

A:Accession: S39792

A:Molecule type: mRNA

A:Residues: 1-160 <12A>

Query Match 5.2%; Score 108; DB 2; Length 160;

Best Local Similarity 24.4%; Pred. No. 0.4;

Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKSEKLRERIKYCCCEQLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVMAQIT 199

Db 28 HNALEKRRDHKIDKSFHSLRDSVPSLOGEKASRAQILDKATEYIOYMRK-----N 78

QY 200 EALQSNMRFCKKQOTPELSLPGTVMAORENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLLEQVRALEKA-RSSAQLQTNPPSSDNLTYNAKGTISAFDGG 137

QY 254 STPDSESLDE 264

Db 138 SDSSESEPEE 148

RESULT 15

I39410

AF-4 protein, splice form 3 - human

N:Alternate names: serine/proline-rich FBL protein

C:Species: Homo sapiens (man)

C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997

C:Accession: I39410

R:Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Cavanaugh, O.; Kamada, N.; Gale, R.P.; L

Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993

A:Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute l

A:Reference number: A47440; MUID:93281633

A:Accession: I39410

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1210 <NAK>

A:Cross-references: GB:I13773; NID:g306446; PID:g306447

C:Comment: This protein is one of several to form chimeric fusion proteins with the t

C:Genetics:

A:Gene: GDB:MUT2; AF-4; FBL

A:Cross-references: GDB:136792; OMIM:159557

A:Map position: 4q21-4q21

C:Keywords: alternative splicing

Query Match 5.2%; Score 108; DB 2; Length 1210;

Best Local Similarity 21.4%; Pred. No. 7.4;

Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCPSNTYKTENAT--GPEELGLPIQ-----RSYSEHGLY-----PPTDLFACSES 114

Db 891 TCGDPPKASASTSKSNKDSIPQRVEGKSGRSSEHSGSGDTANPPV-----PS 944

QY 115 LRANGGLNANSLSEFEKKNKISLH-----SKSEKLRERIKYCCCEQLRTLLPYVK- 166

Db 945 LPNGNS---KPGKQVYFDKQADLHREAKMKQKALMTDRGKAFKLEAVLSIEC 1001

QY 167 -----GKNDASVLEATVDYKYI-----REKISPAVMAQITEALQSN 205

Db 1002 GIATESEQSSKSAVSYSSETVDIKFIMLSKSPDATAPQEKIFAVLCMRQOSILMA 1061

QY 206 MRFCKKQOTPELSLPGTVMAORENSVMSTSPERGLOFLINTCWNGCSTPDSESLDEA 265

Db 1062 MFRCKKD--IAIKYSRTLKHFESSSKVAQAP-----SPCIASGTGPPSPSPSP 1110

QY 266 VRVPSASASENAIGDPYKTHISSAALSLNSLHT--VRYSEKVTSPSYD---ATAVTNQN 318

Db 1111 ASSVGSQSSAGSVSSGVAATISIPVTIQNMTSSYVITISHVLTAFLPLMDQAEALTKN 1169

Search completed: January 19, 2001, 10:46:10

Job time: 32456 sec

Tue Jan 23 10:54:45 2001

us-09-389-000-2.rpr

Page 7

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 03:27:29 ; Search time 62.37 Seconds
(without alignments)
209.702 Million cell updates/sec

Title: US-09-389-000-2

Sequence: 2095 1 MVLRKVPSSLNAELEAIKLT.....MILKAPKDLISKELAMFGF 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115.5	5.5	732	1 HS9A_MOUSE	P07901 mus musculus
2	113.5	5.4	731	1 HS9A_HUMAN	P07900 homo sapien
3	112.5	5.4	732	1 HS9A_PIG	P02705 sus scrofa
4	110	5.3	160	1 MAX_MOUSE	P28574 mus musculus
5	108.5	5.2	728	1 HS9A_CHICK	P11501 gallus galli
6	108	5.2	160	1 MAX_CHICK	P2162 gallus galli
7	108	5.2	160	1 MAX_HUMAN	P25912 homo sapien
8	108	5.2	160	1 MAX_RAT	P2164 rattus norv
9	108	5.2	1210	1 AF4_HUMAN	P21825 homo sapien
10	108	5.2	1653	1 CLH_YEAST	P21137 saccharomyc
11	107.5	5.1	337	1 HAIR_DROME	P14003 drosophila
12	107.5	5.1	1495	1 MAX1_HUMAN	Q13233 homo sapien
13	107	5.1	378	1 HAIR_DROVI	P29303 drosophila
14	106.5	5.1	732	1 HS9A_CRIGR	P46633 cricetus
15	106	5.1	3744	1 YHP9_YEAST	P38811 saccharomyc
16	105.5	5.0	1244	1 SLAI_YEAST	P32790 saccharomyc
17	105	5.0	377	1 Y835_YEAST	Q58245 methanococ
18	104.5	5.0	209	1 MAD4_MOUSE	Q60948 mus musculi
19	104.5	5.0	435	1 DPN_DROME	Q26263 drosophila
20	104.5	5.0	486	1 RTG3_YEAST	P38165 saccharomyc
21	104.5	5.0	725	1 AGAL_YEAST	P23233 saccharomyc
22	104	5.0	209	1 MAD4_HUMAN	Q14582 homo sapien
23	103	4.9	725	1 HS9B_CHICK	Q04619 gallus galli
24	102.5	4.9	1536	1 SIN3_YEAST	P22579 saccharomyc
25	101.5	4.8	163	1 MAX_XENLA	O07016 xenopus lae
26	101.5	4.8	723	1 HS9B_MOUSE	P11469 mus musculi
27	100.5	4.8	486	1 KPR5_YEAST	Q12255 saccharomyc
28	100	4.8	861	1 ORC1_HUMAN	Q13415 homo sapien
29	99.5	4.7	3866	1 HRX_MOUSE	P55200 mus musculi
30	99	4.7	1476	1 AT7A_CRIGR	P49045 cricetus
31	98.5	4.7	726	1 HS9A_BRARE	Q90474 brachydanio
32	98	4.7	901	1 MSH5_YEAST	Q12175 saccharomyc
33	97	4.6	1306	1 MSB2_YEAST	P32334 saccharomyc

34	96.5	4.6	446	1 TFE3_MOUSE	O64092 mus musculus
35	96	4.6	416	1 PAX6_COTJA	P47238 coturnix co
36	96	4.6	422	1 PAX6_HUMAN	P26367 homo sapien
37	96	4.6	422	1 PAX6_MOUSE	P32117 mus musculi
38	96	4.6	422	1 PAX6_RAT	P70601 rattus norv
39	95.5	4.6	1121	1 EXSC_HAEN	P44945 haemophilus
40	95	4.5	363	1 DP3B_AOUAE	O67725 aquifex aeo
41	95	4.5	437	1 PAX6_ORYLA	O73917 oryzae lat
42	95	4.5	666	1 NOD_DROME	P18105 drosophila
43	94.5	4.5	1772	1 MSPI_PLAYO	P13828 plasmodium
44	94	4.5	415	1 PROA_BACSU	P39821 bacillus su
45	94	4.5	575	1 YRAM_HAEN	P45299 haemophilus

ALIGNMENTS

RESULT	ID	HS9A_MOUSE	STANDARD	PRT	732 AA.
AC	P07901	HS9A_MOUSE	01-AUG-1988 (Rel. 08, Created)		
DT	01-FEB-1991	(Rel. 17, Last sequence update)			
DT	15-DEC-1998	(Rel. 37, Last annotation update)			
DE	HSP86	SHOCK PROTEIN HSP 90-ALPHA (HSP 86) (TUMOR SPECIFIC			
DE	HSP86	OR HSP86-1 OR HSP86.			
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BALB/C;				
RX	MEDLINE=89174568; PubMed=2925609;				
RA	Moore S.K., Kozak C., Robinson E.A., Ullrich S.J., Appella E.;				
RT	"Murine 86- and 84-kDa heat shock proteins, cDNA sequences,				
RT	chromosome assignments, and evolutionary origins.";				
RL	J. Biol. Chem. 264:5343-5351(1989).				
RN	[2]				
RP	SEQUENCE OF 5-355 FROM N.A.				
RX	MEDLINE=89232740; PubMed=2469626;				
RA	Hoffmann T., Hovemann B.;				
RT	"Heat-shock proteins, Hsp84 and Hsp86, of mice and men: two related				
RT	genes encode formerly identified tumour-specific transplantation				
RT	antigens.";				
RL	Gene 74:491-501(1988).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92009001; PubMed=1916807;				
RA	Moore S.K., Appella E., Villar C.J., Kozak C.A.;				
RT	"Mapping of the mouse 86-kDa heat-shock protein expressed gene				
RT	(Hsp86-1) on chromosome 12 and related genes on chromosomes 3, 4, 9,				
RT	and 11.";				
RL	Genomics 10:1019-1029(1991).				
RN	[4]				
RP	SEQUENCE OF 459-732 FROM N.A.				
RX	MEDLINE=90033873; PubMed=2806771;				
RA	Legagneux V., Mezger V., Quelard C., Barnier J.V., Bensaude O.,				
RT	Morange M.;				
RT	"High constitutive transcription of Hsp86 gene in murine embryonal				
RT	carcinoma cells.";				
RL	Differentiation 41:42-48(1989).				
RN	[5]				
RP	SEQUENCE OF 1-30.				
RX	MEDLINE=86205848; PubMed=3458168;				
RA	Ullrich S.J., Robinson E.A., Law L.W., Williamson M., Appella E.;				
RT	"A mouse tumor-specific transplantation antigen is a heat shock-				
RT	related protein.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3121-3125(1986).				
RN	[6]				
RP	PHOSPHORYLATION BY DS-DNA KINASE.				
RX	MEDLINE=90008887; PubMed=2507541;				
RA	Lees-Miller S., Anderson C.W.;				

"The human double-stranded DNA-activated protein kinase phosphorylates the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal threonine residues." J. Biol. Chem. 264:17275-17280(1989).

- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY (BY SIMILARITY).

- SUBCELLULAR LOCATION: CYTOPLASMIC.

- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

CC

DR EMBL: J04633; AAA53068.1; -

DR EMBL: M36830; AAA37868.1; ALT_SEQ.

DR EMBL: M57673; AAA37867.1; -

DR EMBL: X16857; CAA34748.1; -

DR PIR: B32848; HHS86.

DR HSSP: P07900; IYES.

DR MGD: MGI:96250; HSP86-1.

DR INTERPRO: IPR001404; -

DR PFAM: PF00183; HSP90; 1.

DR PRINTS: PR00775; HEATSHOCK90.

DR PROSITE: PS00298; HSP90; 1.

DR Chaperone; ATP-binding; Heat shock; Phosphorylation.

KW INIT MET 0

FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).

FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).

FT MOD_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 262 262 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 6 6 T -> A (IN REF. 2).

FT CONFLICT 242 245 MISSING (IN REF. 2).

FT CONFLICT 355 355 R -> K (IN REF. 2).

SEQUENCE 732 AA; 84656 MW; E6345DBEABAC0F1B CRC64;

Query Match 5.5%; Score 115.5; DB 1; Length 732;
Best Local Similarity 22.8%; Pred. No. 0.46;
Matches 87; Conservative 57; Mismatches 126; Indels 111; Gaps 22;

OY 6 VPSASAELEAKLIRFEKK--KNTSLFYFI--PENFGCISGHMDIAL--TEP 57
DB 394 LPLNTSREMLQOSKILIKYRKMLVKCLEFTLEADKENVKRYEQFSKNKILGHEDS 453
OY 58 LTMKMSNVKWTCPSTNTVKTENATGPEELGL--PLORSYSEHLGYPT--DLF 109
DB 454 QNRKRLSELRLYYT-----SASGDEWVSLNDYCTRMKNENKHI-YFITGETKKDY 502
OY 110 ACS---ESIRNGNGLLEL-----NASLSEFEKNKKIS-----LHSSKEKLR 148
DB 503 ANSAVERLRK-HGLEVIYMIPEIDECYQQLKEFGKTLVSTKEGLLPDEEEKKKQ 561
OY 149 ERIKCCCEGLRLPLPYVGRKNDASVLEATYDYKYTEIKI-SPRVAMQITLALOSNR 207
DB 562 EEKKTKFEVLCIKM-----KDLLEKKEKVVVSNRLVTSPPCIATSYGTANNE 611
OY 208 FCKKOOTPIELSLPGTVMAQRENSVSTSPRGLOFLNTGMCSTPDASSIDEAVR 267
DB 612 RIMKQO-----ALRDNSTMGYMAKKHLEI-----NPD--HSITETLR 647
OY 268 VPS-SSASENAIGD---PYKTHISSAALSLN--SLHTVRYYSKY-----TPSYDA 311
DB 648 OKAEADKNDKSVKDLVILLYETALLSSGFSLEDPQTHANRIMTKILGLIGDDPPTVD 707
OY 312 T--ATNTQNIISHLPSAMPV 330
DB 708 TSAAYTEE-----MPPL 719

RESULT 2
HS9A_HUMAN STANDARD; PRT; 731 AA.
AC P07900;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEAT SHOCK PROTEIN HSP 90-ALPHA (HSP 86).
CN HSP90 OR HSPCL OR HSP90A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERIPHERAL BLOOD LYMPHOCYTES;
RX MEDLINE=89386066; PubMed=2780322;
RA Soeda E., Yokoyama K., Yamazaki M., Akaoji K., Miwa T., Imai T.;
RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock protein from human peripheral blood lymphocytes.";
RL Nucleic Acids Res. 17:7108-7108(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91242090; PubMed=1368637;
RA Yamazaki M., Tashiro H., Yokoyama K., Soeda E.;
RT "Molecular cloning of cDNA encoding a human heat-shock protein whose expression is induced by adenovirus type 12 ELA in HeLa cells.";
RL Agric. Biol. Chem. 54:3163-3170(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RX MEDLINE=89343979; PubMed=2527334;
RA Hickey E., Brandon S.E., Smale G., Lloyd D., Weber L.A.;
RT "Sequence and regulation of a gene encoding a human 89-kilodalton heat shock protein.";
RL Mol. Cell. Biol. 9:2615-2626(1989).
RN [4]
RP SEQUENCE OF 1-311 FROM N.A.
RX MEDLINE=88056312; PubMed=2445630;
RA Hoffmann T., Hoyemann B.;
RT "Cloning and nucleotide sequence of the murine hsp84 cDNA and chromosome assignment of related sequences.";
RL Gene 56:29-40(1987).
RN [5]
RP SEQUENCE OF 1-311 FROM N.A.
RX MEDLINE=90076956; PubMed=2591742;
RA Walter T., Drabent B., Krebs H., Tomalak M., Heiss S.,
RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding gene.";
RL Gene 83:105-115(1989).
RN [6]
RP SEQUENCE OF 1-20. AND PHOSPHORYLATION.
RX MEDLINE=89123325; PubMed=2492519;
RA Lees-Miller S., Anderson C.W.;
RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated in vitro by casein kinase II.";
RL J. Biol. Chem. 264:2431-2437(1989).
RN [7]
RP PHOSPHORYLATION BY DS-DNA KINASE.
RX MEDLINE=90008887; PubMed=2507541;
RA Lees-Miller S., Anderson C.W.;
RT "The human double-stranded DNA-activated protein kinase phosphorylates the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal threonine residues.";
RL J. Biol. Chem. 264:17275-17280(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.
RX MEDLINE=97262065; PubMed=9108479;
RA Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,
RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent.";

```

RL Cell 89:239-250(1997).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.
RA MEDLINE=99034582; PubMed=9817749;
RA Obermann W.M., Sondermann H., Russo A.A., Pavletich N.P., Hartl F.U.;
RT "In vivo function of Hsp90 is dependent on ATP binding and ATP
RT hydrolysis."
RL J. Cell Biol. 143:901-910(1998).
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC -!- (BY SIMILARITY)
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15183; CA33259.1; -
DR EMBL: X07270; CA30255.1; -
DR EMBL: M27024; AA63194.1; -
DR EMBL: M30626; AA36023.1; -
DR PIR: A32319; HHU086.
DR PIR: B31420; B31420.
DR PIR: J00724; J00724.
DR PDB: 1YER; 22-APR-98.
DR PDB: 1YES; 22-APR-98.
DR PDB: 1YET; 22-APR-98.
DR PDB: 1BYO; 28-OCT-98.
DR TRANSPAC: T00992; -
DR MIM: 140571; -
DR INTERPRO: IPR001404; -
DR PIRAM: PF00183; HSP90.1.
DR PRINTS: PR00775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90.1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation; 3D-structure.
FT INIT_MET 0
FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).
FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).
FT MOD_RES 230 230 PHOSPHORYLATION.
FT MOD_RES 262 262 PHOSPHORYLATION.
FT CONFLICT 62 62 T -> S (IN REF. 3, 4 AND 5).
FT SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;
SQ
Query Match 5.4%; Score 113.5; DB 1; Length 731;
Best Local Similarity 21.8%; Pred. No. 0.65;
Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;
OY 6 VPSSLSAELEAIKILRFGK---KNTSLFVFII--PENFGCISGHGMDIAL--TEP 57
DB 393 LPLNISREMLQOSKILKYLKRLVKKCLELTELADKENVKRYEFSKNIKLGIHDS 452
OY 58 LTMKMSNVKWTCPSTVYKTNATGPEELG---PLORSYSEHLGYFTDL----- 108
DB 453 QNRKLISELLRYTT-----SASGDEWVSLKRYCTRMKENQKHYYITGETKDOYA 502
OY 109 -FACSESILRNGLEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
DB 503 NSAFYERLRK-HGLEVIYMIPIDEYCYOOLKEFGKILVSTKRGLELPEDEEEKKOE 561
OY 130 RIKYCCEDLRTLLPYVKGKKNDAVLEATVDYVYIREKI--SPAVMAQITELALSNMF 208
DB 562 EKKTFEENLCKIM-----KDLTEKKYKVVVSNRLVSPCCIVSTGWTANMR 611
OY 209 CKKOOTPIELSLPGTYMAORENSVMSTYSPERGLOFLTMCNNGCSTPAESLDEAVV 268
DB 612 IKAQ-----ALRDNSTMGYMAKKHLEI-----NPD--HSIETTLKO 647
OY 269 PS-SSASENAIGD---PYKTHISSAALSLN--SLHTVYYS--KVTPSYDATAVTNOMI 319

```

```

DB 648 KAEADKNKSVADLYLLYETALLSSGFSLEDDPQTHANRYIMKILGIDEDDEPTADT 707
OY 320 SIHLPSAMPV 330
DB 708 SAAVTEMPPL 718
RESULT 3
HS9A_PIG STANDARD: PRT; 732 AA.
ID HS9A_PIG
AC 002705;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HEAT SHOCK PROTEIN HSP 90-ALPHA (HSP 86).
GN HS9A OR HSP90A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Huang H.W., Jian S.C., Huang H.C., Lee W.C.;
RT "Complete nucleotide sequence of a porcine cDNA for 90-kDa heat
RT shock protein."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC -!- (BY SIMILARITY)
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U94395; AAC48718.1; -
DR INTERPRO: IPR001404; -
DR PIRAM: PF00183; HSP90.1.
DR PRINTS: PR00775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90.1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT INIT_MET 0
FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 230 230 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 262 262 PHOSPHORYLATION (BY DS-DNA KINASE).
FT SEQUENCE 732 AA; 84643 MW; 759850062F4E5490 CRC64;
SQ
Query Match 5.4%; Score 112.5; DB 1; Length 732;
Best Local Similarity 21.8%; Pred. No. 0.77;
Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;
OY 6 VPSSLSAELEAIKILRFGK---KNTSLFVFII--BENFGCISGHGMDIAL--TEP 57
DB 394 LPLNISREMLQOSKILKYLKRLVKKCLELTELADKENVKRYEFSKNIKLGIHDS 453
OY 58 LTMKMSNVKWTCPSTVYKTNATGPEELG---PLORSYSEHLGYFTDL----- 108
DB 454 QNRKLISELLRYTT-----SASGDEWVSLDKYCTRMKENQKHYYITGETKDOYA 503
OY 109 -FACSESILRNGLEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
DB 504 NSAFYERLRK-HGLEVIYMIPIDEYCYOOLKEFGKILVSTKRGLELPEDEEEKKOE 562
OY 150 RIKYCCEDLRTLLPYVKGKKNDAVLEATVDYVYIREKI--SPAVMAQITELALSNMF 208

```

```

DB 563 EKTKEPKNLCKIM-----KDLIEKVEKVVNSRLVTSQCCVSTGYGTANMER 612
OY 209 CKKQGPPIELSLPQTVAAQRENSVSYSPERGLOFLNTQWNGCSPDAESSIDEAVRY 268
DB 613 IMKRA-----ALRDNSTMGMAKKHLEI-----NPD--HSITTEFLQ 648
OY 269 PS-SSASENAIGD---PYTHISSALSLN--SLHVRVRY--KYPPSDATAVTQNT 319
DB 649 KAADKNDKSVKDLVILYETALLSGFSEDPQTHANRYMIKIGLIGIDEDDPTADDS 708
OY 320 SIHPSAMPV 330
DB 709 SAAVTEMPPL 719

RESULT 4
MAX_MOUSE STANDARD; PRT; 160 AA.
AC P28574;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAX PROTEIN (MYN PROTEIN) (MYC-BINDING NOVEL HLH/LZ PROTEIN).
GN MAX OR MYN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
RN Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RP SEQUENCE FROM N.A.
RX MEDLINE-91208676; PubMed-1840505;
RA Prendergast G.C., Lave D., Ziff E.B.;
RT "Association of Myc, the murine homolog of max, with c-Myc stimulates
RT methylation sensitive DNA binding and ras cotransformation.";
RL Cell 65:395-407(1991).
RN [2]
RP STRUCTURE BY NMR OF 70-102 IN COMPLEX WITH MYC.
RX MEDLINE-98347001; PubMed-9680483;
RA Lavigne P., Crump M.P., Gagne S.M., Hodges R.S., Kay C.M., Sykes B.D.;
RT "Insights into the mechanism of heterodimerization from the 1h-NMR
RT solution structure of the c-Myc-Max heterodimeric leucine zipper.";
RL J. Mol. Biol. 281:165-181(1998).
CC -1- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC DNA-
CC BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES THE CORE
CC SEQUENCE 5'-CAC(GA)TG-3'. THE MYC-MAX COMPLEX IS A TRANSCRIPTIONAL
CC ACTIVATOR. WHEREAS THE MAD-MAX COMPLEX IS A REPRESSOR. CPG
CC METHYLATION OF THE RECOGNITION SITE GREATLY INHIBITS DNA BINDING.
CC SUGGESTING THAT DNA METHYLATION MAY REGULATE THE C-MYC/MAX COMPLEX
CC IN VIVO.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. THE SHORT FORMS DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION (BY SIMILARITY).
CC -1- INDUCTION: IN 3T3 FIBROBLASTS BY SERUM.
CC -1- PTM: PHOSPHORYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63903; AAA39797.1; -
DR PIR: A38488; A38488.
DR PDB: 1A93; 18-NOV-98.
DR RDB: 2A93; 23-FEB-99.
DR TRANSFAC: T00523; -

```

```

DR MGI:96921; MAX.
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFM: PFM0010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation; 3D-structure.
FT DNA_BIND 24 36 BASIC DOMAIN
FT DOMAIN 39 75 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 81 102 LEUCINE-ZIPPER (BY SIMILARITY).
FT VARSPLIC 13 21 MISSING (IN SHORT ISOFORM) (BY
FT SEQUENCE 160 AA; 18197 MW; CBA5E1894E984F56 CRC64;
SQ

Query Match 5.38; Score 110; DB 1; Length 160;
Best Local Similarity 24.4%; Pred. No. 0.14;
Matches 32; Conservative 27; Mismatches 56; Indels 16; Gaps 3;

OY 140 HSKERLRREIRKYCCQQLRTLLPYKGRKNDASVLEATVDVYKIRKISPAVAQIT 199
DB 28 HNALEKRRDHINDSHSLDSVPSLQGERASQAQLDKATEIYQYRRK-----N 78
OY 200 EALQSNMRECKKOQTPIELSLPQTVAAQRENSVSYSPERGLOFLTN-----TCWNGC 253
DB 79 DTHQDDIDDLKRONALLDQGVRALEKA-RSSAQLQNTYPSDMSLYTNKKGITSAFDGG 137
OY 254 STPDASSLDE 264
DB 138 SDDSSESEPE 148

RESULT 5
HSA9_CHICK STANDARD; PRT; 728 AA.
ID HSA9_CHICK
AC P11501;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT SHOCK PROTEIN HSP 90-ALPHA.
GN HSPCA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-STOMACH; AND GIZZARD;
RX MEDLINE-89165846; PubMed-2923621;
RA Binart N., Chambraud B., Dumas B., Rowlands D.A., Bigogne C.,
RA Levin J.M., Garnier J., Baulieu E.E., Catelli M.G.;
RT "The cDNA-derived amino acid sequence of chick heat shock protein Mr
RT 90,000 (HSP 90) reveals a 'DNA like' structure: potential site of
RT interaction with steroid receptors.";
RL Biochem. Biophys. Res. Commun. 159:140-147(1989).
RN [2]
RP SEQUENCE OF 1-85 FROM N.A.
RX MEDLINE-89345085; PubMed-2762125;
RA Vourc H.C., Binart N., Chambraud B., David J.P., Jerome V.,
RA Balleu E.E., Catelli M.G.;
RT "Isolation and functional analysis of chicken 90-kDa heat shock
RT protein gene promoter.";
RL Nucleic Acids Res. 17:5259-5272(1989).
CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07265; CA30251.1; -
DR EMBL: X15028; CA33132.1; -
DR PIR: S10880; HRC90.
DR HSSP: P07900; IYES.
DR INTERPRO: IPR001404; -
DR PFM: PF00183; HSP90; 1.
DR PRINTS: PRO0775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90; 1.
KM Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT MOD_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 259 259 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 728 AA; 84059 MW; 6EC5660FAD5006CE CRC64;

Query Match 5.2%; Score 108.5; DB 1; Length 728;
Best Local Similarity 21.6%; Pred. No. 1.5;
Matches 80; Conservative 58; Mismatches 142; Indels 91; Gaps 18;

OY 6 VPSSASAELEAKILIRGK---KNTSLFVFII--PENFGCISGHGMDIAL---TEP 57
DB 390 LPLNTSRMLQOSKILKYLKLVKKCELTETELAEDEKENVKFEQTSKNIKLGHDS 449
OY 58 LTMKMSVVKVWYTCPSNTVKTENATGPDELGL---PLQSYSEHLGYPTDL----- 108
DB 450 QNRKLTSLRLRYT-----SASGDEWVSLDKYCRMKENKHVYIIGERKDOYA 499
OY 109 -FACSESLRNGDEL-----NASEFEKRNKIS-----LHSSKRLRE 149
DB 500 NSAFERLRK-HGLEVIYVIEPIDEYCVQOLKEFEKGLVSTKGLLEPEDEKKKQOE 558
OY 150 RIKYCCEDRLTLPYVGRKNDASVLEATVDYVYIREKI-SPAVMAQITFALQSNMF 208
DB 559 EKKAFENLCKTM-----KYLEKKYKVVVNSPLVSPCCITVTSTGTANNER 608
OY 209 CKKOQPIELSLPTVMAQRENSVMSTYSPERGLOFLNTCNGCSTPDASSLDAVAV 268
DB 609 IMKAQ-----ALRNSTMGYMAKKHLEI-----NPD--HSIETLRLQ 644
OY 269 PS-SSASENAIGD---PYKTHISSAALSLN--SLHTVAYIS--KVTBSYDNTANTNONT 319
DB 645 KAEADKNKSVKDYLLIYETALLSSGFLEDPQTHANRYIMIKLGLIDEDDTAAEBA 704
OY 320 SIHLSPMPV 330
DB 705 SPATVEEMPPL 715

RESULT 6
MAX_CHICK STANDARD: PRT: 160 AA.
AC P52162;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAX PROTEIN.
GN Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FIBROBLAST;
RX MEDLINE-94119609; PubMed-8290277;
RA Solleberger K.G., Kao T.L., Taparowsky E.J.;
RT "Structural analysis of the chicken max gene.";
RL Oncogene 9:661-664(1994).
CC -!- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC[CA]TG-3'. THE MYC-MAX COMPLEX IS A

```

```

CC TRANSSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR (BY SIMILARITY).
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. THE SHORT FORM DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION.
CC -!- PTM: PHOSPHORYLATED (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12469; AAA16834.1; -
DR HSSP: P25912; 1AN2.
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation.
FT DNA_BIND 24 36 BASIC DOMAIN.
FT DOMAIN 39 75 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 81 102 LEUCINE-ZIPPER (BY SIMILARITY).
FT VARSPIC 13 21 MISSING (IN SHORT ISOFORM) (BY
FT SIMILARITY).
SQ SEQUENCE 160 AA; 18184 MW; F09B16D1AB379A7E CRC64;

Query Match 5.2%; Score 108; DB 1; Length 160;
Best Local Similarity 23.7%; Pred. No. 0.2;
Matches 31; Conservative 26; Mismatches 58; Indels 16; Gaps 3;

OY 140 HSSKELRERIRKYCCEDRLTLPYVGRKNDASVLEATVDYVYIREKISPVMQIT 199
DB 28 HNALEKRRDRHKDSFSLRDSVPSLGEKASRAQIDLKATEIYIYMRK-----N 78
OY 200 EALQSNMFRCKKQOTPIELSLPTVMAQRENSVMSTYSPERGLOFLN-----TCWNGC 253
DB 79 HTHOODIDDLKRNALLLEQVRALEKARSSAQLQANTY-PADSSLYTPKGSTISAPFG 137
OY 254 STPDASSLDE 264
DB 138 SDSSSDSEPE 148

RESULT 7
MAX_HUMAN STANDARD: PRT: 160 AA.
AC P25912; P52163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAX PROTEIN.
GN Homo sapiens (Human), and Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-HUMAN;
RX MEDLINE-91173288; PubMed-2006410;
RA Blackwood E.M., Eisenman R.N.;
RT "Max: a helix-loop-helix zipper protein that forms a
RT sequence-specific DNA-binding complex with Myc.";
RL Science 251:1211-1217(1991).

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D14447; BAA03337.1; -
 DR EMBL: D14448; BAA03338.1; -
 DR HSSP: P25912; IAN2
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR003013; -
 DR PFM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.
 KW Nuclear protein; DNA-binding; transcription regulation;
 KW alternative splicing; phosphorylation.
 FT DNA BIND 24 36 BASIC DOMAIN.
 FT DOMAIN 39 75 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 81 102 LEUCINE-ZIPPER (BY SIMILARITY).
 FT VARSPIC 13 21 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 160 AA; 18272 MW; 86CB1A137727A57A CRC64;

Query Match 5.2%; Score 108; DB 1; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.2;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKERKREKIKKCCOEOLRTLLPVKGRKNDASVLEATVDYVYKREKISPAVMAQIT 199
 Db 28 HNALEKRRRRDRIKSFHSLRDSVPSLOGEKASRAQIIDKATIEYIOMRK-----N 78

QY 200 EALOSNRFECKKQJOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 79 HTHOODIDDLKKNALAEQVRALEKA-RSSAQLOTNYPSSDNSLYTNAGKGTISAFDGG 137

QY 254 STPDSESLIDE 264
 Db 138 SDSSSESEPEP 148

RESULT 9
 AF4_HUMAN STANDARD; PRT; 1210 AA.

AC P51825;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE AF-4 PROTEIN (FEEL PROTEIN).
 GN MLT2 OR AF4 OR FEU.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93281633; PubMed-8506309;
 RA Nakamura T., Alder H., Gu Y., Prasad R., Canaan O., Kamada N.,
 RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
 RA Canaan E.;
 RA "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
 RT acute leukemia share sequence homology and/or common motifs";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93184301; PubMed-8443374;
 RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
 RA Cleary M.L.;
 RA "A serine/proline-rich protein is fused to HRX in t(4;11) acute
 RT leukemias";
 RT Blood 81:1124-1131(1993).
 RL -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DISEASE: INVOLVED IN A T(4;11)(Q21;Q23) CHROMOSOMAL TRANSLOCATION
 CC IN ACUTE LEUKEMIAS CAUSING FUSION TO THE MLT/HRX GENE PRODUCT
 CC WHICH CONTAINS DNA-BINDING MOTIFS RESULTING IN A ROGUE ACTIVATOR

CC PROTEIN.
 CC -1- SIMILARITY: TO IAF4 AND FMR2 (OX19).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: L13773; AAA58360.1; -
 DR EMBL: L25050; AAA36642.1; -
 DR MIM: 159557; -
 KW Nuclear protein; Chromosomal translocation; proto-oncogene.
 FT DOMAIN 483 492 POLY-SER.
 FT DOMAIN 835 843 POLY-SER.
 FT DOMAIN 866 869 POLY-PRO.
 FT DOMAIN 871 874 POLY-SER.
 FT CONFLICT 46 46 K -> R (IN REF. 2).
 FT CONFLICT 624 624 E -> G (IN REF. 2).
 FT CONFLICT 899 905 SASSTKS -> VPAVPRV (IN REF. 2).
 FT CONFLICT 928 929 EH -> AD (IN REF. 2).
 FT CONFLICT 999 999 I -> N (IN REF. 2).
 FT CONFLICT 1096 1096 A -> AR (IN REF. 2).
 FT CONFLICT 1140 1140 N -> I (IN REF. 2).
 STNCVTLANSLVDLYHTRGROLOELTKTP -> RQ
 MCAPWPTAVMWTCITHDVFESSYKN (IN REF. 2).
 SQ SEQUENCE 1210 AA; 131421 MW; F0E334DF8CE2FF04 CRC64;

Query Match 5.2%; Score 108; DB 1; Length 1210;
 Best Local Similarity 21.4%; Pred. No. 3.3;
 Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCSNVTYKTENAT--GPEELGLPQ-----RSYSEHLGY-----PPTDLPACES 114
 Db 891 TCGDDPKRSASSTKSNHKSISIPKORVEKGRSSSRSSSGDGTANPFPV-----PS 944

QY 115 LRNGNGLNLNSLSEPEKKNKISLH-----SKKELRERKIKYCCOJRTLLPYVK 166
 Db 945 LPNGNS--KPKQPVAFDQKQADLHMRKAKKOKKALMDRGAFAKYLEAVLSFEC 1001

QY 167 -----GRKNDASVLEATVDYVYKTI-----REKISPAVMAQITREALQSN 205
 Db 1002 GIARESSQSKSAVSYSEVVDLKFIMLSKFSDDATAPQEKIFPAVLCRCOSILMA 1061

QY 206 MRECKKQOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLTNCGWNGCSTPDSESLIDEA 265
 Db 1062 MFRCKKD---IAIKYSRTLNHFESSKVAQAP-----SPCIASGTGFPSPPLSPMPSP 1110

QY 266 VRVSSASASEMAIDDPKTHSSAALSLNSLHT--VRYKSYVTSTYD---ATAVTNQN 318
 Db 1111 ASSVSQSSAGVSGVAVATISPTVITQNTSSYVITSHVLTAFDLMEQAEALTRKN 1169

RESULT 10
 CLH_YEAST STANDARD; PRT; 1653 AA.

AC P22137;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE CLATHRIN HEAVY CHAIN.
 GN CHC1 OR YGL206C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRATN-8288C;
 RX MEDLINE-91093343; PubMed-1898742;
 RA Lemmon S.K., Pellicena-Palle A., Conley K., Freund C.L.;

DR EMBL: X15904; CAA34018.1; -
 DR EMBL: X15905; CAA34019.1; -
 DR PIR: S06956; S06956.
 DR TRANSFAC: T00345; -
 DR FLYBASE: FBgn0001168; h.
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR003015; -
 DR PFAM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
 DR Nuclear protein: Developmental protein; Pair-rule protein;
 DR DNA-binding: Transcription regulation; Repressor.
 FT DNA_BIND 32 44
 FT DOMAIN 45 89
 FT DOMAIN 149 157
 FT DOMAIN 222 237
 FT DOMAIN 241 250
 FT VARIANT 292 292
 FT DOMAIN 334 337
 SO SEQUENCE 337 AA; 36995 MW; 6D2ECAF7FD56C0B CRC64;
 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 S -> P.
 Query Match 5.1%; Score 107.5; DB 1; Length 337;
 Best Local Similarity 23.7%; Pred. No. 0.62; Mismatches 84; Indels 55; Gaps 10;
 Matches 53; Conservative 32; Mismatches 84; Indels 55; Gaps 10;
 QY 128 SEFEKNNKISLHSSKEKLRERIKYCCOQLRTL-----PYVGRKNDASVLEATVD 181
 DB 30 SDRSKNPI-----MEKRRRRINNCLNELKTLIDATKDPARHSKLEKADILEKTVK 83
 QY 182 YVYIYREK---ISPAVMAQTLEALQSNMRCCKKQQTPIELSLGTVMAQENSVMSYSP 238
 DB 84 HLOELDROQAAMQADPKLVNFKAGFADCVNEVS---RFGDIEPAQR----- 129
 QY 239 ERGLQELTNTCMNGCST-----PDSESLDEAVVPSSA---SENIGDPYKTHIS 287
 DB 130 RRLQHLHN-CINGVKTELHQORQOQOSIHQMLPSPSSPEQDSQCAAPYLFGLQ 168
 QY 288 SAALSL--NSLHTVRYYSKVPSTYDATAVTQNTNISHLPSAMP 328
 DB 189 QFASGYFLPMGMQVI-----PTKLPMGSLALVLPQSLP 221
 RESULT 12
 M3L_HUMAN STANDARD; PRT; 1495 AA.
 AC 013233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE MITOGEN-ACTIVATED PROTEIN KINASE 1 (EC 2.7.1.-)
 DE (MAPK/ERK KINASE KINASE 1) (MEK KINASE 1) (FRAGMENT).
 GN MAP3K1 OR MAPKKK1 OR MEKK1 OR MEKK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN (1)
 RN SEQUENCE FROM N.A.
 RA MEDLINE-9902611; PubMed-9808624;
 RA Xia Y., Wu Z., Su B., Murray B., Karin M.;
 RT "JNK1 organizes a MAP kinase module through specific and sequential
 RT interactions with upstream and downstream components mediated by its
 RT amino-terminal extension.";
 RL Genes Dev. 12:3369-3381(1998).
 RN (2)
 RN SEQUENCE OF 1221-1257 FROM N.A.
 RC TISSUE-LEUCOCYTE;
 RX MEDLINE-96169565; PubMed-8597633;
 RA Vink B.S., Kay E.S., Fiedorek F.T.;
 RT "Mapping of the MEK Kinase gene (MeKk) to mouse chromosome 13 and
 RT human chromosome 5.";
 RL Mamm. Genome 6:782-783(1995).
 CC -1- FUNCTION: CAN PHOSPHORYLATE AND ACTIVATE MAPK1 AND MAPK2
 CC (MEK1/MEK2) WHICH LEADS TO PHOSPHORYLATION OF MAP KINASES. IT IS

CC ALSO A HIGHLY EFFICIENT ACTIVATOR OF THE JNK CASCADE.
 CC -1- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE KINASE KINASE SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF042838; AAC97073.1; -
 CC EMBL: U29671; AAB05828.1; -
 CC DR SWISS-2DPAGE; Q13233; HUMAN.
 CC MIM: 600982; -
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR002290; -
 DR PFAM: PF00069; pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Phosphorylation.
 FT NON_TER 1 1
 FT DOMAIN 6 10 POLY-GLY.
 FT DOMAIN 17 22 POLY-ALA.
 FT DOMAIN 403 412 POLY-SER.
 FT DOMAIN 430 442 PH.
 FT DOMAIN 633 740 PH.
 FT DOMAIN 925 932 POLY-THR.
 FT DOMAIN 1165 1170 POLY-GLU.
 FT DOMAIN 1199 1202 POLY-ILE.
 FT DOMAIN 1226 1481 PROTEIN_KINASE.
 FT NP_BIND 1232 1239 ATP (BY SIMILARITY).
 FT BINDING 1255 1255 ATP (BY SIMILARITY).
 FT ACT_SITE 1352 1352 BY SIMILARITY.
 SO SEQUENCE 1495 AA; 163653 MW; 2D6ECEB52E7AFCAFB CRC64;
 Query Match 5.1%; Score 107.5; DB 1; Length 1495;
 Best Local Similarity 24.1%; Pred. No. 4.9; Mismatches 95; Indels 65; Gaps 10;
 Matches 62; Conservative 35; Mismatches 95; Indels 65; Gaps 10;
 QY 129 EFEKNNKISL-----HSSKEKLRERIKYCCOQLRTLPPYVGRKNDASVLEA 178
 DB 769 EIRYKLLSLTLTALQSIDNSHMGRLSR-RYLSARAVTVVPHVFSKLEMLSVSSV 827
 QY 179 TVDYVYIREKISPAVMAQTLEALQSNMRCCKKQQTPIELSLGTVMAQENSV----- 232
 DB 828 SHFTMRRLRLMAYADVEIEALIQ-----LGVEDTLQROOHNSFCRHLFP 873
 QY 233 MSTYSEBRLQF-----LTYTCMNGCSTPDSESLDEAVR-----PSSASAEKAI--- 278
 DB 874 TTIWKPORTVPLECTVLEKGTGKLCATKLSASSEDISERLARISVGPSSSTTTTTE 933
 QY 279 -----GDPYKTHISSAALSNS--LHTVRYYSKVPSTYDATAVTQNTNISHLPSAMP 933
 DB 934 QPKPMVOTKCRPHSQCLNSPLSHSQLMFPLALSTSSSTPSVPAQTANDVSKHRLOGFI 993
 QY 320 SIHLPSAMPVS--SFSL 335
 DB 994 PCRIIPASPTQRRKFSL 1010
 RESULT 13
 HAIR_DPROVI STANDARD; PRT; 378 AA.
 ID HAIR_DPROVI
 AC P29303;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HAIRY PROTEIN.
 GN H.
 OS Drosophila virilis (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92269819; PubMed=1588951;
 RA Mainwring S.M., Ish-Horowicz D.;
 RT "Point mutations in the Drosophila hairy gene demonstrate *in vivo*
 RT requirements for basic, helix-loop-helix, and WRPW domains.";
 RT Mol. Cell. Biol. 12:2475-2483(1992).
 CC -1- FUNCTION: PAIR-ROLE PROTEIN THAT REGULATES EMBRYONIC SEGMENTATION
 CC AND ADULT BRISTLE PATTERNING. TRANSCRIPTIONAL REPRESSOR OF GENES
 CC THAT REQUIRE A BHLH PROTEIN FOR THEIR TRANSCRIPTION (EG. THE FUSHI
 CC TARAZU GENE)
 CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
 CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
 CC HELIX-INTERERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
 CC RATHER THAN THE CANONICAL E-BOX (CANNG).
 CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
 CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
 CC HAIRY-RELATED PROTEINS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M87885; AAA28602.1; -
 CC DR TRANSFAC: T01643; -
 CC DR FLIRBASE: FBgn0013115; DVLrNh.
 CC DR INTERPRO: IPR001092; -
 CC DR INTERPRO: IPR003015; -
 CC DR PFAH: PF00010; HLH: 1.
 CC DR PROSITE: PS00038; HELIX_LOOP_HELIX: 1.
 CC KW Nuclear protein; Developmental protein; Pair-rule protein;
 CC DNA-binding; Transcription regulation; Repressor.
 CC FT DOMAIN 22 28 POLY-GLN.
 CC FT DNA_BIND 38 50 BASIC DOMAIN.
 CC FT DOMAIN 51 95 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 CC FT DOMAIN 249 267 GLN-RICH.
 CC FT DOMAIN 269 283 POLY-ALA.
 CC FT DOMAIN 375 378 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 CC FT DOMAIN 378 378 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 CC SQ SEQUENCE 378 AA; 41383 MW; D3C2D565AA57FCAE CRC64;

 Query Match 5.1%; Score 107; DB 1; Length 378;
 Best Local Similarity 23.3%; Pred. No. 0.79; Indels 66; Gaps 11;
 Matches 56; Conservative 36; Mismatches 82; Indels 66; Gaps 11;
 OY 128 SEFEKNNKISILHSSKEKLRERIKYCCEDLRLT-----PYVGRKNDAAVLEATVD 181
 DB 36 SDRSKNKP-----MEKRRRARINNNCLNELTKLIDATKKDPAHSHKLEKADILEKTVK 89
 OY 182 YVYIREK---ISPAYMAQITLALQSMWRCKKQKQITLIELSPCTVAAQENSMVSTYSP 238
 DB 90 HLOELROQAAMQAAADPKYIINKFAGFADCAEVN---RFPGLDSTOR----- 135
 OY 239 ERLQGLTNTCNWGCSP-----DASSIDEAVRPSSASNAIGDPYKTISS 288
 DB 136 RRLLOHLN--CINGVKTWHLHGHGHALQAQGLHAAQVYLPSPSPSP--EBSVTVYAA 192

OY 289 AA---LSLNSLHIVRY-----SKVPSYDAVAVTNQNSIHLPSAMP 328
 DB 193 SGNSSNNNTTAPYLFGLQIQNANGYFLPNGQVLP-----FKLPNGSLALVLPQSLP 248

 RESULT 14
 HS9A_CRIGR STANDARD; PRT; 732 AA.
 ID HS9A_CRIGR
 AC P46833;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP 90-ALPHA (HSP 86).
 GN HS9A OR HSP90A.
 OS Cricetus griseus (Chinese hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Cricetus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chen M.S.M.C., Laszlo A.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: I33676; AAA36992.1; -
 CC DR HSSP: P07900; IYES.
 CC DR INTERPRO: IPR001404; -
 CC DR PFAH: PF00183; HSP90; 1.
 CC DR PRINTS: PR00775; HEATSHOCK90.
 CC DR PROSITE: PS00298; HSP90; 1.
 CC KW Chaperone; ATP-binding; Heat shock; Phosphorylation.
 CC FT INIT_MET 0 4
 CC FT MOD_RES 4 4
 CC FT MOD_RES 6 6
 CC FT MOD_RES 230 230
 CC FT MOD_RES 262 262
 CC FT SEQUENCE 732 AA; 84717 MW; AAF118902PAB3402 CRC64;

 Query Match 5.1%; Score 106.5; DB 1; Length 732;
 Best Local Similarity 22.6%; Pred. No. 2.1;
 Matches 86; Conservative 58; Mismatches 126; Indels 111; Gaps 22;
 OY 6 VPSSLSAELEAIKILIRGKK---KTHSLFVFIIT--PENFGICISGHMDIAL---TEP 57
 DB 394 LPLNISREILLOOSKIKVIRKKNLVKRCLELFHLEADKENVKRYEYQFSKNIKLGIHDS 453
 OY 58 LIMEKSNVYKVTGCPSTVNTKENTATGPELGL-----PLQRSYSHLGYFP--DLE 109
 DB 454 QNRKLSLELRYIT-----SASGDEWYSLKRYCTRMKNQKHI-YFITGERKDDV 502
 OY 110 ACS---ESLNGNGLEL-----NASISEFEKNNKIS-----LLHSSKELLR 148
 DB 503 ANSAFVFLRK-HGLVIVMIEPIDEYVQOLKEFGKTLVSYTKEGLPEDEEEKKRO 561
 OY 149 ERIKYCCEDLRLTLPYVGRKNDAAVLEATVDYVYIIEKI-SPAYMAQITLALQSNMR 207
 DB 562 EEKKTFFENLCKIM-----KDILKVKYKVVVSNRLVYSPCCIVTSTYGTWATNE 611
 OY 208 FCKKQQTPIELSLPGTVAAQRENSVSTYSPERGLQFLNTCNWGCSTPDASSLDEAVR 267

```

Db      612 RIRAO-----ALRDNSTMGYMAARKHLEI-----NPD--HSIETLR 647
QY      268 VPS-SSASENAIGD---PYKTHISSAALSLN--SLHTVRYSKY-----TPSDA 311
Db      648 OKAEADKDKSVKDLVILLIETALSSGFSLEDPOTANRIRYKIKLGLIDEDDPYDD 707
QY      312 T--AVTNONISIHLPSPMPV 330
Db      708 TSAVTEE-----MPL 719

```

```

RESULT 15
YHP9_YEAST
ID YHP9_YEAST STANDARD; PRT: 3744 AA.
AC P38811:

```

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 433.2 KDA PROTEIN IN HXT5-NRK1 INTERGENIC REGION.
GN YHR099W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE-94378003; PubMed-8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO S.POMBE SPAC1F5.11C.
CC -1- SIMILARITY: BELONGS TO THE P13/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; U00060; AAB68923.1; -
CC
CC PIR; S46715; S46715.
CC
CC SGD; S0001141; YHR099W.
CC
CC Hypothetical protein; Transferase; Kinase.
CC
CC DOMAIN 3414 3744 P13K/P14K.
CC
CC SEQUENCE 3744 AA; 433171 MW; AE358676F5D5777 CRC64;

```

```

Query Match      5.1%; Score 106; DB 1; Length 3744;
Best Local Similarity 23.0%; Pred. No. 22;
Matches 58; Conservative 40; Mismatches 114; Indels 40; Gaps 9;

```

```

QY      123 LNASLSEFEKNKKISLHSSKEKLRRERIKYCC-----EQLRTLLPYVGRKNDAA 173
Db      1821 LKSYLVEDKKRPMELHKNLTKNSNATLAVDLDHDLFPPELLQLSAIFIKA---DPE 1877
QY      174 SVLEATVDYVYIRE--KISPAVMAQITEALQSNMRPKKQOTPIELSLPQTVAQRENS 231
Db      1878 IIAEIKKDIKFCWNFKILEDTLIKQSAIYVTS--YFISKDFPIKVTQYVVALNRSSH 1935
QY      232 VMSYSPERGLOFLT---NTCMNGCSTPDASSLDEAVRVPSSASASENAIGDPYKTHISS 288
Db      1936 VEARYLVKQSDVLTLPVYHEHMAAGTPTDITWINVKRYVENSSQNNIL---YQFLISH 1992

```

```

QY      289 AALSLNS-----LHTVRYSKVTPSDATAVTNONISIHLPSPMPVSSSLGTALLGMA 343
Db      1993 PDLFPNSRDLFISNIHMMNKI-----TFMSNSNSDSH-----TLAIDLASLIYWE 2039
QY      344 RRAHPIPYCNS 355
Db      2040 NKTLEITVNMNT 2051

```

```

Search completed: January 19, 2001, 10:52:08
Job time: 26679 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 19, 2001, 02:57:44 ; Search time 65.98 Seconds
(without alignments)
719.448 Million cell updates/sec

Title: US-09-389-000-2
Perfect score: 2095
Sequence: 1 MWLKVPSLSAELEAIKIL.....MIKAPPKDISKELEWGF 405

Scoring table: BLOSUM62
Gapop 10.50, Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1699	81.1	425	4 Q9NX45	Q9NX45 homo sapien
2	146	7.0	685	10 Q41875	Q41875 zea mays (m
3	121	5.8	367	10 Q65552	Q65552 arabidopsis
4	119	5.7	452	4 Q94771	Q94771 homo sapien
5	119	5.7	452	4 Q9U149	Q9U149 homo sapien
6	114	5.4	1788	13 Q9Y6K0	Q9Y6K0 oreochromis
7	113.5	5.4	539	4 Q75322	Q75322 homo sapien
8	113	5.4	311	10 Q9S0S3	Q9S0S3 arabidopsis
9	112	5.3	350	10 Q9ZUG9	Q9ZUG9 arabidopsis
10	112	5.3	359	10 Q9M270	Q9M270 arabidopsis
11	111	5.3	254	10 Q9XEF0	Q9XEF0 arabidopsis
12	110	5.3	662	10 Q9ZPY8	Q9ZPY8 arabidopsis
13	109	5.2	339	11 Q9QUS4	Q9QUS4 mus musculu
14	109	5.2	1708	4 Q94895	Q94895 homo sapien
15	109	5.2	1773	4 Q9Y6B8	Q9Y6B8 homo sapien
16	109	5.2	1922	4 Q43355	Q43355 homo sapien
17	109	5.2	3595	4 Q9UCH3	Q9UCH3 homo sapien
18	109	5.2	3908	4 Q9UC04	Q9UC04 homo sapien
19	109	5.2	3911	4 Q99996	Q99996 homo sapien

20	107.5	5.1	337	5 Q9VSN8	Q9VSN8 drosophila
21	107	5.1	415	10 Q23192	Q23192 arabidopsis
22	106.5	5.1	1006	2 P74686	P74686 synechocyst
23	105.5	5.0	297	10 Q9LSQ3	Q9LSQ3 arabidopsis
24	105.5	5.0	446	4 Q9NVH3	Q9NVH3 homo sapien
25	105.5	5.0	796	4 Q9P217	Q9P217 homo sapien
26	105	5.0	460	5 P91751	P91751 loligo opal
27	105	5.0	4588	5 Q94709	Q94709 parametium
28	104.5	5.0	204	13 Q9W6H1	Q9W6H1 xenopus lae
29	104.5	5.0	435	5 Q9V384	Q9V384 drosophila
30	104	5.0	313	5 Q18054	Q18054 caenorhabdi
31	103	4.9	297	10 Q9SAB6	Q9SAB6 arabidopsis
32	103	4.9	423	10 Q9SKX6	Q9SKX6 arabidopsis
33	103	4.9	2627	4 Q99973	Q99973 homo sapien
34	103	4.9	3899	4 Q9Y6X2	Q9Y6X2 homo sapien
35	102.5	4.9	314	10 Q9U072	Q9U072 arabidopsis
36	102.5	4.9	524	10 Q9SBC5	Q9SBC5 arabidopsis
37	102.5	4.9	848	2 Q9RAT4	Q9RAT4 lactococcus
38	102.5	4.9	1492	11 P70705	P70705 rattus norv
39	102	4.9	310	10 Q22768	Q22768 arabidopsis
40	102	4.9	337	4 Q9UBP5	Q9UBP5 h basic-hel
41	102	4.9	859	5 Q21401	Q21401 caenorhabdi
42	101.5	4.8	329	10 Q81306	Q81306 arabidopsis
43	101.5	4.8	416	8 Q78234	Q78234 chrysosplen
44	101.5	4.8	416	8 Q78237	Q78237 chrysosplen
45	101.5	4.8	979	5 Q9VEN2	Q9VEN2 drosophila

ALIGNMENTS

RESULT 1
Q9NX45 PRELIMINARY; PRT; 425 AA.
ID Q9NX45:
AC Q9NX45:
DT 01-OCT-2000 (TIREMBLrel. 15, Created)
DT 01-OCT-2000 (TIREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TIREMBLrel. 15, Last annotation update)
DE CINA FLJ20449 FIS, CLONE KAT05575.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matanabe K., Kunagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000456; BAA91175.1; -
SQ SEQUENCE 425 AA; 46922 MW; 777E57231F0D6BDD CRC64;

Query Match	81.1%	Score 1699;	DB 4;	Length 425;
Best Local Similarity	89.4%	Pred. No. 4.1e-126;		
Matches 338;	Conservative 5;	Mismatches 11;	Indels 24;	Gaps 4;
QY 1	MWLVKVPSSLSAELEAIKILRFQKKNTSLFVFIIPENFKGISGHGMDIALTEPLTM 60			
DB 67	MWLVKVPSSLSAELEAIKILRFQKKNTSLFVFIIPENFKGISGHGMDIALTEPLTM 126			
QY 61	EKMSNVKWTTCPSNVKTEENATGPELGLPILORSYSEHLGYPTDLFACSESLRNGC 120			
DB 127	EKMSNVKWTTCPSNVKTEENATGPELGLPILORSYSEHLGYPTDLFACSESLRNGC 186			
QY 121	LELNASISEFEKKKIKSLHSSKKELRRERIKYCCEDLRLLPYVKKRKNDASVLEATV 180			
DB 187	LELNASISEFEKKKIKSLHSSKKELRRERIKYCCEDLRLLPYVKKRKNDASVLEATV 246			
QY 181	DYVYTIKESIPAYMAOITFALOSNMRFCKKQOPIELSLPGTYMAORENSVMTYSR 240			
DB 247	DYVYTIKESIPAYMAOITFALOSNMRFCKKQOPIELSLPGTYMAORENSVMTYSR 306			

QY 241 GLOFLTNMGNGSTPDASSIDEAVRPSSASANAIDGPYKTHISSAALSTNSHTYR 300
 DB 307 GLOFLANTWNGSTPDASSIDEAVRPSSASANAIDGPYKTHISSAALSTNSHTYR 366
 QY 301 YSKVTPSYDATAVTNQNISIHLPSPVPSSFLTGALGMARALHPTVCNS-FGRI 359
 DB 367 YSKVTPSYDATAVTNQNISIHLPSPVPSSFLTGALGMARALHPTVCNS-FGRI 411
 QY 360 KST---CLKFTLSTTYWA 374
 DB 412 CTHPNCLO-----QFWA 424

RESULT 2
 ID 041875 PRELIMINARY: PRT: 685 AA.
 AC 041875;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE REPRESSOR-LIKE PROTEIN (INI).
 GN INI.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W22;
 RA Burr F.A.; Burr B.; Scheffler B.E.; Biewitt M.; Wienand U.; Matz E.C.;
 RL Plant Cell 0.0-0(0).
 DR EMBL: U57899; AAB03841.1;
 DR MENDEL: 11294; Zeama.1300; 11294.
 DR INTERPRO: IPR001092;
 DR INTERPRO: IPR003015;
 DR PFM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN.1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN.1.
 SQ SEQUENCE 685 AA; 74698 MW; FDC67F8E95F21F79 CRC64;

Query Match 7.0%; Score 146; DB 10; Length 685;
 Best Local Similarity 23.5%; Pred. No. 0.0027;
 Matches 73; Conservative 46; Mismatches 111; Indels 80; Gaps 14;

QY 89 LGPLPQRSSEHIGYPTDLPFACSESLRNGNGLNLAS-LSSEPKKKTSLHSSSEKR 147
 DB 421 LGAPSSSSSHRSHRG---EVQSSSPRRDDEGTSRSRGPVPSQTELSASHVLEKERR 476
 QY 148 RRIKRCCEBLPLPYVGRKNDASVLEATVDYKIRKISPAVMAQITFALQSNR 207
 DB 477 RKLNEGFMALSLVPEV--TKMDRASILGDTIEYVKQLRRRI-----QELSSRR 525
 QY 208 FCKKOQPIELSLPGTVMAQRENSVMSTYSPERGQ----FLTNMGNGSTPDASSL 262
 DB 526 LVGSNQK-----TTMAQPPPPAAS-TEERGRQTSGGYLARAAGSRAEASGNS 576
 QY 263 DEAVRPSSASAE---NAIGPYKTHISSAALSLNSLHTVRYKSVTSDATAVT 315
 DB 577 NGEEPPAASATDTEVOVSIIG-----SPALLELCRPREGLLHVMAQ-----L 623
 QY 316 NONISIHLPSPVPSSFLTGALGMAR-----RALHI-----PTVCN 354
 DB 624 HOELRELEITS---VOASSAGVLLAKLRAKKEVHGRRSSITEVRAIHLIVSSDWICE 679
 QY 355 SFGRITSTCL 364
 DB 680 -----KNPCL 684

RESULT 3
 ID 065552 PRELIMINARY: PRT: 367 AA.

AC 065552;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 39.6 KDA PROTEIN.
 GN F6118.110 OR AT4G30980.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M.; Benes V.; Rechmann S.; Borkova D.; Ansoerge W.; Hohelsel J.;
 RA Mewes H.W.; Mayer K.; Schueller C.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V.; Rechmann S.; Borkova D.; Ansoerge W.; Mewes H.W.; Lemcke K.;
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL022198; CAA18195.1;
 DR EMBL: AL161578; CAB79816.1;
 DR MENDEL: 29120; Arabid.3244; 29120.
 DR INTERPRO: IPR001092;
 DR PFM: PF00010; HLH; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 39613 MW; 80A5E4E8F4ECE6DA CRC64;

Query Match 5.8%; Score 121; DB 10; Length 367;
 Best Local Similarity 24.9%; Pred. No. 0.11;
 Matches 45; Conservative 36; Mismatches 82; Indels 18; Gaps 5;

QY 140 HSKKELRERKRYCCEDQLTLLPYVGRKNDASVLEATVDYKIRKISPAVMAQIT 199
 DB 198 HSIARLRERKRIARMKSLQELVP--NGKKTDKAMSLDEITDIYVFLQYKVLMSRLG 255
 QY 200 EALQSNMRCKKQPIELSLPGTVMAQRENSVMSTYSPERG--LQFLTNMGNGSTPD 257
 DB 256 GAASASSQISSEDAGSHENTSSSGEAKMTEHQAALMEDGNSAQYLGK--GLCLMPI 313
 QY 258 AESSIDEAVRPSSASANAIDGPYKTHISSAALSL---NSLHTVRYKSVTSDATAVT 314
 DB 314 SLATTISTATCPSSRSPFVKDGVPLSPNLSTIYVANGSSSLVTVK-----DAPSV 364
 QY 315 T 315
 DB 365 S 365

RESULT 4
 ID 094771 PRELIMINARY: PRT: 452 AA.
 AC 094771;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE TRANSCRIPTION FACTOR-LIKE 5.
 GN TCFU5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Maruyama O.; Katagiri T.; Nishimori H.; Miki Y.; Ueno A.; Nakamura Y.;
 RT "Cloning, mapping, and genomic organization of THH1, a novel human
 testis-specific gene containing a basic Helix-loop-helix motif."

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96267138; PubMed=9602032;
 RA Schweinfest C.W., Graber M.W., Henderson K.W., Papas T.S., Baron P.L.,
 Watson D.K.;
 RT "Cloning and sequence analysis of Hsp90alpha DeltaN, a new member of
 the hsp90 gene family."
 RL Biochim. Biophys. Acta 1398:16-24(1998).
 DR EMBL: AF028832; AAC25497.1; --
 DR INTERPRO: IPR001404; --
 DR PFAM: PF00183; HSP90; 1.
 SQ SEQUENCE 539 AA; 63251 MW; F10A801D002B6547 CRC64;

Query Match 5.4%; Score 113.5; DB 4; Length 539;
 Best Local Similarity 21.8%; Pred. No. 0.72;
 Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;

QY 6 VPSSIAEELAIKLRFGK--KNTSLFEVFI--PENFGGCGIGHMDIAL--TPP 57
 DB 201 LPLNLSREMLQOSKILKVRKNLVKCLFTELAECKENIKFYQFSKNIKLGIHEDS 260
 QY 58 LTMKSNVVKWTCTPSTVKTENATGPELGL---PLQSYSEHLGYFPFDL----- 108
 DB 261 QNRKKLSLRLRYT-----SASGDEWVSLKDYCRMENQKHIIYITGETKDQYA 310
 QY 109 -FACSESLNGLNGLEL-----NASLSEFEKNKKIS-----LLHSKEKLRE 149
 DB 311 NSAFYERLRK-HGLEVIYMEPIDEXVOOLKEFEKGKTLVSYTKEGLELPEDEEEKKOE 369
 QY 150 RIKYCCOELRTLPPYKGRKNDASVLEATVDYVKIRREKI-SPVMAQITEALQSNMF 208
 DB 370 EKKTFENICKIM-----KDLKKVKEVYVSNLVSPCCIVSTIGWYANMER 419
 QY 209 CKQQTPIELSLPGVMAQRENSVNSTYSPERGLOFLNTCWNCGSTPDASSLDEAVRY 268
 DB 420 IMKAQ-----ALRDNSTMGYMAKKHLEI-----NPD--HSIITELQ 455
 QY 266 PS-SSASENAID---PKTHISSAALSUN--SLHTVRYYS--KTPSYDATATYANQVI 319
 DB 456 KMEADKNDKSVADLVLLYETALLSSGFSLEDPQTHANRIYMKILGIDDDPTADDT 515
 QY 320 SIHLSPMPV 330
 DB 516 SAAVTEEMPL 526

RESULT 8
 Q9SOS3 PRELIMINARY; PRT; 311 AA.
 AC Q9SOS3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE DNA-BINDING PROTEIN.
 GN T6L1.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome I BAC T6L1 genomic sequence."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC011665; AAF07356.1; --
 DR INTERPRO: IPR001092; --
 DR PFAM: PF00010; HLH; 1.

KW DNA-binding.
 SQ SEQUENCE 311 AA; 34487 MW; CC72A040A1A2E427 CRC64;

Query Match 5.4%; Score 113; DB 10; Length 311;
 Best Local Similarity 20.1%; Pred. No. 0.37;
 Matches 57; Conservative 46; Mismatches 125; Indels 56; Gaps 8;

QY 99 EHLGFPFDLPACSSSLNGLNGLELNASLSEFEKNKKISLHSSKEKLRERIKCCQL 158
 DB 9 EEDYGEEDFNSKRGPSSNNTVHNSRDSKENDKASAIRKSHSYEORRRKINERFQIL 68
 QY 159 RFLPYVGRKNDASVLEATVDYVKIRKIS-----PVMQITEALQSNMFC 209
 DB 69 RELINSE-QKRDYASFLLEVYDIYQYQAEKYQKEGSPKSGSEPTLTPRRNHRVQ 127
 QY 210 KQQTPIELS-----LPGTVMAQRENSVST---YSPERGLOFLNTCWNCGSTPD 257
 DB 128 SLGNHPVAINNNGSGPIFPCKF--EDNTVSTPAAIAEQPIESDKARAITGISI-E 183
 QY 258 AESSLDEAVRP-----SSASFNAGDPYKTHISSAALSLSLHTVRYYS 303
 DB 184 SQPELDDKGLPLQILPVMQEQANECPATSDGQSDNLVIEGTTISSAVSHELLS 243
 QY 304 KTPSYDATAV-----TNQISIHLPAMPVVS 331
 DB 244 SITQALQNAIGDLSQAKLSVQIDLGKRAHQGLTHEPSSKNPLS 287

RESULT 9
 Q9ZUG9 PRELIMINARY; PRT; 350 AA.
 AC Q9ZUG9;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE F27D4.17 PROTEIN.
 GN F27D4.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F27D4 genomic sequence."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005967; AAD03387.1; --
 DR MENDEL: 38842; Arabidopsis3244.38842.
 DR INTERPRO: IPR001092; --
 DR PFAM: PF00010; HLH; 1.
 SQ SEQUENCE 350 AA; 36522 MW; 8438D16596C636B9 CRC64;

Query Match 5.3%; Score 112; DB 10; Length 350;
 Best Local Similarity 37.5%; Pred. No. 0.52;
 Matches 27; Conservative 16; Mismatches 17; Indels 12; Gaps 2;

QY 140 HSSKEKLRERIKCCOELRTLPPYKGRKNDASVLEATVDYVKIRREKI----- 190
 DB 149 HSIARLRERIERAKMKALQELVP--NGNKTDKASMLDEIIDYKFLQYKVLMSRLG 206
 QY 191 -SPVMAQITEA 201
 DB 207 GAASVSQISEA 218

RESULT 10
 Q9M270 PRELIMINARY; PRT; 359 AA.
 ID Q9M270

```

AC 09M270;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 41.2 KDA PROTEIN.
GN F21P14.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choiane N., Robert C., Brottier P., Mincker P., Catolico L.,
RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Meyer K.F.X., Queller F., Salanoubat M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL138642; CAB71902.1; -.
KW Hypothetical protein.
SQ SEQUENCE 359 AA; 41158 MW; BCATFAT95EEDBDBA CRC64;

Query Match
Best Local Similarity 22.3%; Score 112; DB 10; Length 359;
Matches 65; Conservative 4; Mismatches 107; Indels 76; Gaps 12;

QY 82 NATGPEELGPL-----QSYSEHLGYFPDPLFACSESLRNGLELNASISEFNKKKI 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 NMTLPSTSTSLSAHSRKRKRNHL--LPQEMTREKRKRKRKTKPSKNNEEI---EMQRI 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 SLHSSSEKRLREKIRKCCEDRLTLPLPYVGRKNDASVLEATVDYKRIEKSIPVMA 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 N-HIAEERRRRRMMNHINSLRALLPSTYIQRDQASIVGAINYK-----VLE 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 QTEALSNMFKCKKQOTPIELSLPGTMAQRENSVMSTYSPERGLOFLTNTCNGCSTP 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 228 QIIOSLESQR-----TQOQNSSEV-----VENALNHL-----SGTSSN 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 DAESSLDEAVRPSSAS-----ENAIQDPYKTHISSAALSLSLH--TVRY 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 DLMTTLEDQCTIPKIEATVIONHVSLLKQVCEKKGQLKGIISLEKLTALVHLNITSS 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 YSKVTPSYDATAVTNNQNIHSLPSAMPVSSFSGLTALGMARALHPTV 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 HSSVSYSEFNKQMEDE---CDLESA-----DETTAAVHRIFDIPPTI 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
Q9XEFO PRELIMINARY; PRT; 254 AA.
AC Q9XEFO;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DE 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 28.4 KDA PROTEIN.
GN T07M07.8 OR A12640200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MDLLINE-99225673; Pubmed-10207155;
RA Wang M.L., Belmonte S., Kim U., Dolan M., Morris J.W., Goodman H.M.;
RT "A cluster of ABA-regulated genes on Arabidopsis thaliana BAC
T07M07."
RL Genome Res. 9:325-333(1999).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayen L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."
RL Nature 402:761-768(1999).
DR EMBL: AF085279; AAD25935.1; -.
DR EMBL: AC018721; AAF18734.1; -.
DR INTERPRO: IPR001092; -.
DR PFM: PF00010; HLH; 1.
KW Hypothetical protein.
SQ SEQUENCE 254 AA; 28414 MW; EE3BCE353EE52619 CRC64;

Query Match
Best Local Similarity 22.2%; Score 111; DB 10; Length 254;
Matches 69; Conservative 47; Mismatches 109; Indels 86; Gaps 13;

QY 79 KTEATGPEELGPILOMSYE-----HLCYFPDPLFACSESLRNGLELNASISEFEK 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9 KWSPTSTPYVSWSLQESSSDSMNRPNLGF-----SSSSFGN-PPADCVGGRK 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 133 NKRSLSHSEKRLRRIRIKYCCEDRLTLPLPYVGRKNDASVLEATVDYKRIEKSIP 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 60 AESLSRSHRLAEKRRDRINSHLTALKIYP--NSDKLDLAALAIYIEQVKEKQAA- 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 AYVAQITAEALQSNMFKCKKQOTPIELSLPGTMAQRENSVMSTYSPERGLOFLTNT 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 -----ESPFIQDLP-----TEADEVTYQPTISDFESNTIIL 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 -----CNGCSTPDAESSLDEAVRPSSASENAIGDPYTHISSAALSLSLHVTYYS 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 FKASFC---CE--DQPEAISEIIRVLKIQLET-----IQAEIISVGRMRNIFIL 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 KVPESYDAT--AVTNNQNIHSLPSAMPVSSFSGLTALGMARALHPTVCSFGRKS 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 KDSNCNETTINIASAKALKOSLCSALNRTSSSTTTS-----SYC-----RIRS 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 362 TCLKFTLSTY 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 KRWKFLSSHX 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
Q9ZPY8 PRELIMINARY; PRT; 662 AA.
AC Q9ZPY8;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE F13A10.4 PROTEIN.
GN F13A10.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
RA Carreira A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC F13A10 genomic sequence."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006418; AAD20162.1; -.
DR INTERPRO: IPR001092; -.

```

DR INTERPRO: IPR003015; -
 DR PFAM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 SO SEQUENCE 662 AA; 74346 MW; A6C78EC03AAB53F CRC64;

Query Match 5.3%; Score 110; DB 10; Length 662;
 Best Local Similarity 21.0%; Pred. No. 1.8; Mismatches 154; Indels 100; Gaps 17;
 Matches 83; Conservative 58;

3 LKVPSSSAEELAKLI-RFGKKNTSLFVETIP- - - - -NFGCIS--GKG 49
 :
 :
 207 IVVPTDAGVLLGWSVSPENIGLVQALFMRVTPVWVTSNTMTGIIHKLFGOD 266
 :
 :
 50 MDIALTEPTLMKMSVYVWYWTCPSTNTKTNANGPEELGPIQR - - - - - 95
 :
 :
 267 LSGAAVYPKKLVRRRLDERFT--POSMEGYNNKGP-TFGTTPQDDVYKLENNVVD 323
 :
 :
 96 --SYSEHLGYPTDFA- - - - -CSE- - - - -SLRNGCLEINASLSEFEK 134
 :
 :
 324 NNNYQTGFASVVAASNPSTNQEKSECTEKRPVSLLAGI - - - - -VGVDEKRR 380
 :
 :
 135 KTS- - - - -LHSSKEKLRREIKKCCQQLTLPYVGRKNDASVLEAVDYK 184
 :
 :
 381 KGRKPRANGREPLNHVEAEORREKLNRFYALRSVFNIS--KKDKASLLGDAISYIK 438
 :
 :
 185 YREKISPAVMAQITEALOSNMRFCKKQOTPIELSLPGVMAQRENSVSTSP- - - - - 238
 :
 :
 439 ELQEVK--IMED--ERVGTDLSLSESNITTYEESPEVDIQMNEVYVRAVSLDISHA 494
 :
 :
 239 ERGLQFLNTCNGCSTPDASSLDEAVRPSSASENAGDPYKTHISSALINSLSLT 298
 :
 :
 495 SRIIQAMR--SNVSLMAKLSLAEDTMFHFVIRKSNNGSDPLTFREKLIAMF- - - - - 544
 :
 :
 299 VVYKSVTPSYDATAVTNONISIHLPASMPVSSP 333
 :
 :
 545 - - - - -YPTSTSTOP- - - - -PLPSSSSQVSGF 565
 :
 :
 RESULT 13
 090US4 PRELIMINARY: PRT: 339 AA.
 ID 090US4
 AC 090US4
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE BASIC HELIX-LOOP-HELIX FACTOR 1 (BASIC-HELIX-LOOP-HELIX PROTEIN) (HES-RELATED REPRESSOR PROTEIN 1 HERP1).
 DE CHE1 OR HRT2 OR HEY2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chin M.T., Maemura K., Fukumoto S., Jain M., Layne M., Matanabe M., Hsieh C.-M., Lee M.-E.;
 RT "Cardiovascular basic helix-loop-helix factor 1, a novel transcriptional repressor expressed preferentially in the developing and adult cardiovascular system.";
 RL J. Biol. Chem. 0:0-0(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nakagawa O., Nakagawa M., Richardson J.A., Olson E.N., Srivastava D.;
 RT "HRT1, HRT2 and HRT3: A New Subclass of bHLH Transcription Factors Marking Specific Cardiac, Somitic and Pharyngeal Arch Segments.";
 RL Dev. Biol. 216:72-84(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Steidl C., Leimeister C., Klamt B., Maier M., Nanda I., Schmid M., Gessler M.;
 RT "Characterisation of the human and mouse Hey genes: cloning, mapping and mutation screening of a new bHLH gene family.";

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Leimeister C., Externbrink A., Klamt B., Gessler M.;
 RT "Hey genes: a novel subfamily of hairy- and enhancer of split related genes specifically expressed during mouse embryogenesis.";
 RL Mech. Dev. 85:173-177(1999).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Iso T., Sartorelli V., Polzat C., Chen T., Sucov H.M., Chung G., Wu H.-Y., Kedes L., Hamamori Y.;
 RT "The HERP Repressors, Novel Partners for HES/hairy in Notch Signaling.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF173902; AAF20174.1; -
 DR EMBL: AF172287; AAF14546.1; -
 DR EMBL: AJ271867; CAB71346.1; -
 DR EMBL: AF232240; AAF37298.1; -
 DR INTERPRO: IPR000104; -
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR001993; -
 DR INTERPRO: IPR003015; -
 DR PFAM: PF00010; HLH; 1.
 DR PRINTS: PR00308; ANTFREEZE1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
 SO SEQUENCE 339 AA; 35873 MW; 105BBA06DE50444 CRC64;

Query Match 5.2%; Score 109; DB 11; Length 339;
 Best Local Similarity 21.7%; Pred. No. 0.86;
 Matches 61; Conservative 42; Mismatches 108; Indels 70; Gaps 9;

87 BELGLPQRSYSEHLGYPTDFAFCSESLRNGCLEINASLSEFEKKNKISLHSSKEKL 146
 :
 :
 16 ETIDGSENNTPGH- - - - -ATSSVMRSNSETTTSQIMARKRGIT - - - - -EKR 59
 :
 :
 147 RERIKYCCEQLRTLLPYVGRKNDASVLEAVDYKVIKISPAVM--AQIT 199
 :
 :
 60 RDRINNLSLRLRVPFAFEKQSAKLEKAEILLQMTYDHLKMLQATGKGYFDAHALAT 119
 :
 :
 200 EALOSNMRFCK--KQOTPIELSLPGVMAQRENSVSTSPERGLQFLNTCNGCST 255
 :
 :
 120 DFMSGFECLETVARYLSVEGLDPSDPLRVLRVLSHSTCASOR - - - - - 164
 :
 :
 256 PDASSLDEAVRPSSASENAGDPYK-- - - - -THISSALINSLSHTVRYKVTPEYD 310
 :
 :
 165 - - - - -EAAVMTSSMAHHHPLPHHMAAFHLLPTALLQPNGLHT--SESTPCRL 212
 :
 :
 311 ATAATNONISIHLPASMPVSSFSIGTALLGMARALHPT 351
 :
 :
 213 STS- - - - -SEVSAHGSAALTITFAHADSAALMPS 242
 :
 :
 RESULT 14
 094895 PRELIMINARY: PRT: 1708 AA.
 ID 094895
 AC 094895
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE KIAA0803 PROTEIN (FRAGMENT).
 GN KIAA0803.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUB-BRAIN;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 DR EMBL: AB018346; BAA34523.1; -
 FT NON_TER 1
 SQ SEQUENCE 1708 AA; 197004 MW; A4232490DCB1E229 CRC64;

Query Match 5.2%; Score 109; DB 4; Length 1708;
 Best Local Similarity 20.6%; Pred. No. 7.9; Mismatches 94; Indels 54; Gaps 9;
 Matches 50; Conservative 45;

QY 13 EELEAI-----KLIRFGKK--KNTHSLFVFIIIPENFKGICSGHMDIALTEPLTME 61
 DB 148 EEIQLENEVIEIKQDELANIGQKTSMAHSL-----SEADSLKHQLDVYIAEKLALE 200
 QY 62 KMSNVVKYITTPCSNTVKTE---NATGPEELGLPLQRSYSEHLGTFP----- 105
 DB 201 QOEVANEMETFMKNVLTETNFKMNQLTQELPSLKRRESVEKIOSIPENSVNVAIDHLS 260
 QY 106 -----TDLFACSESLRNGNGLEINASLSEFEKKKISILHSSKEKLRRE-----RIK 152
 DB 261 KDKPELEVLTEDAKS---LENQYTFKSFEEGKSGITINLETRLQLLESTVSAKDEL 317
 QY 153 YCCEDLRTLPLPYVKGKNDASVLEATVDYKYIREKISPAVMAOI-TEALOSNMRFCKK 211
 DB 318 CCYKQIKDM---QEOGOFETEMLOKKIYVLOKIVEEKVAAALVSOILEAVOEYAKFCOD 374
 QY 212 QQT 214
 DB 375 NOT 377

RESULT 15

QY6B8 PRELIMINARY: PRT: 1773 AA.

AC QY6B8:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE SPICE VARIANT AKAP350 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
 RT "AKAP350: A multiply spliced family of proteins with centrosomal
 RT association."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF091711; AAD39719.1; -
 FT NON_TER 1
 SQ SEQUENCE 1773 AA; 204414 MW; 092A3174725D9262 CRC64;

Query Match 5.2%; Score 109; DB 4; Length 1773;
 Best Local Similarity 20.6%; Pred. No. 8.4; Mismatches 94; Indels 54; Gaps 9;
 Matches 50; Conservative 45;

QY 13 EELEAI-----KLIRFGKK--KNTHSLFVFIIIPENFKGICSGHMDIALTEPLTME 61
 DB 195 EEIQLENEVIEIKQDELANIGQKTSMAHSL-----SEADSLKHQLDVYIAEKLALE 247
 QY 62 KMSNVVKYITTPCSNTVKTE---NATGPEELGLPLQRSYSEHLGTFP----- 105
 DB 248 QOEVANEMETFMKNVLTETNFKMNQLTQELPSLKRRESVEKIOSIPENSVNVAIDHLS 307
 QY 106 -----TDLFACSESLRNGNGLEINASLSEFEKKKISILHSSKEKLRRE-----RIK 152
 DB 308 KDKPELEVLTEDAKS---LENQYTFKSFEEGKSGITINLETRLQLLESTVSAKDEL 364

QY 153 YCCEDLRTLPLPYVKGKNDASVLEATVDYKYIREKISPAVMAOI-TEALOSNMRFCKK 211
 DB 365 CCYKQIKDM---QEOGOFETEMLOKKIYVLOKIVEEKVAAALVSOILEAVOEYAKFCOD 421
 QY 212 QQT 214
 DB 422 NOT 424

Search completed: January 19, 2001, 10:50:14
 Job time: 28350 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 23:07:12 ; Search time 2530.03 Seconds
(without alignments)
4304.523 Million cell updates/sec

Title: US-09-389-000-1

Perfect score: 2128
Sequence: 1 gaccggggggcggttggggt.....ccacattatcgaataaaaaa 2128

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank1:
1: gb_ba1:
2: gb_ba2:
3: gb_om:
4: gb_ov:
5: gb_ph:
6: gb_p11:
7: gb_p12:
8: gb_p1:
9: gb_p12:
10: gb_p13:
11: gb_p1:
12: gb_sy:
13: gb_un:
14: em_fun:
15: em_hum1:
16: em_hum2:
17: em_in:
18: em_om:
19: em_or:
20: em_ov:
21: em_pat:
22: em_ph:
23: em_pl:
24: em_ro:
25: em_sts:
26: em_sy:
27: em_un:
28: em_v1:
29: gb_htg1:
30: gb_htg2:
31: gb_in1:
32: gb_in2:
33: em_ba1:
34: em_ba2:
35: em_hum3:
36: em_hum4:
37: gb_p14:
38: gb_htg3:
39: gb_htg4:
40: gb_htg5:
41: gb_htg6:
42: gb_htg7:
43: em_htg1:

44: em_htg2:
45: em_htg3:
46: em_hum5:
47: gb_p13:
48: gb_p15:
49: gb_htg8:
50: gb_htg9:
51: gb_htg10:
52: gb_htg11:
53: gb_htg12:
54: gb_htg13:
55: gb_htg14:
56: gb_in3:
57: gb_htg15:
58: gb_htg16:
59: gb_htg17:
60: em_htg4:
61: em_htg5:
62: em_htg6:
63: em_htg7:
64: em_hum6:
65: gb_htg18:
66: gb_htg19:
67: gb_htg20:
68: gb_htg21:
69: gb_htg22:
70: gb_htg23:
71: gb_v11:
72: gb_v12:
73: gb_ba3:
74: em_htg8:
75: em_htg9:
76: em_htg10:
77: gb_p16:
78: gb_p17:
79: gb_sts1:
80: gb_sts2:
81: gb_pat1:
82: gb_pat2:
83: em_htg9:
84: gb_htg24:
85: gb_p18:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528.6	71.8	2147	37	AK000456
2	337.4	15.9	158888	67	AL160392
3	202	9.3	153964	37	AL139377
4	58.8	2.8	7218	81	166494
5	57	2.7	318	77	HS47C6R
6	48	2.3	2069	21	E10125
7	48	2.3	3399	21	E10126
8	45	2.1	132513	8	AC004520
9	45	2.1	188563	50	AC022764
10	44	2.1	201010	54	AC034122
11	43.8	2.1	132491	53	AC026784
12	43.8	2.1	162162	39	AC012432
13	43.8	2.1	162397	39	AC012536
14	43.8	2.1	174057	58	AC073127
15	43.6	2.0	9985	56	PCU43145
16	43.2	2.0	204207	50	AC023147
17	42.8	2.0	10030	1	AE001134
18	42.8	2.0	141948	29	AC008413
19	42.8	2.0	181195	49	AC021230
20	42.8	2.0	197748	69	AL390840
21	42.4	2.0	226798	50	AC022235
					Mus muscu

Db 54781 ACAATGGCCCTGACCTACACTCTTTCTCAAAAATATTCACAAATTATGAA 54722
OY 2124 aaa 2126
Db 54721 AAA 54719

RESULT 3
AL139377/c
LOCUS AL139377 153964 bp DNA PRI 16-SEP-2000
DEFINITION Human DNA sequence from clone RP11-251J8 on chromosome 13, complete
sequence.
ACCESSION AL139377
VERSION AL139377.8 GI:10185469
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 153964)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Sep 18, 2000 this sequence version replaced gi:9943981.
requests: clonerequests@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-251J8 is from the library RP11-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://dacpac.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-251J8 The true
left end of clone RP11-22001 is at 76095 in this sequence. The true
right end of clone RP11-121N13 is at 46542 in this sequence.

FEATURES
source
1..153964
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-251J8"
/clone_11b="RP11-11.1"
complement(4394..4902)
/note="match: GSS: Em:AQ566052"
4913..5434
/note="match: GSS: Em:B56355"
20504..21023
/note="match: GSS: Em:AQ565583"
complement(28424..28906)
/note="match: GSS: Em:AQ519597"
complement(37716..38198)
/note="match: GSS: Em:AQ726251"
49445..49809
/note="match: GSS: Em:AQ017175"

misc_feature complement(53217..53653)
/note="match: GSS: Em:AQ527490"
misc_feature complement(58957..59671)
/note="match: GSS: Em:AQ375951"
66026..66528
/note="match: GSS: Em:AQ036199"
complement(69728..70139)
/note="match: GSS: Em:AQ218718"
72359..72935
/note="match: GSS: Em:AQ386749"
74084..74298
/note="match: GSS: Em:AQ08740"
74131..74254
/note="match: GSS: Em:AQ079168"
complement(78982..79522)
/note="match: GSS: Em:AQ699184"
complement(79085..79496)
/note="match: GSS: Em:AQ267170"
79965..80602
/note="match: GSS: Em:AQ353722"
81464..81824
/note="match: GSS: Em:AQ027170"
88495..88841
/note="match: GSS: Em:AQ088142"
complement(102691..103082)
/note="match: GSS: Em:AQ435358"
complement(102692..103043)
/note="match: GSS: Em:AQ088271"
complement(105513..105949)
/note="match: GSS: Em:AQ434114"
105972..106395
/note="match: GSS: Em:AQ402870"
106283..106790
/note="match: GSS: Em:AQ282655"
114150..114790
/note="match: GSS: Em:AQ317196"
complement(121221..121649)
/note="match: GSS: Em:B56354"
complement(121255..121609)
/note="match: GSS: Em:AQ466633"
123153..123847
/note="match: GSS: Em:AQ527227"
132103..132500
/note="match: GSS: Em:B55736"
132118..132493
/note="match: GSS: Em:AQ021461"
complement(132120..132493)
/note="match: GSS: Em:AQ344420"
132181..132416
/note="match: GSS: Em:A49391"
complement(132230..132493)
/note="match: GSS: Em:B94003"
132260..132541
/note="match: GSS: Em:AQ627818"
complement(132296..132496)
/note="match: GSS: Em:AQ617736"
complement(132323..132495)
/note="match: GSS: Em:AQ021494"
complement(132355..132502)
/note="match: GSS: Em:AQ034473"
complement(132358..132495)
/note="match: GSS: Em:AQ314706"
complement(132379..132499)
/note="match: GSS: Em:AQ313377 Em:AQ553345"
133586..133962
/note="match: GSS: Em:AQ207068"
147570..147573
/note="152 element removed from here."
BASE COUNT 48904 a 31432 c 28348 g 45280 t
ORIGIN

Query Match 9.5% Score 202; DB 37; Length 153964;

```

Best Local Similarity 100.0%; Pred. No. 3,3e-38;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaccggggggggttggttaccgctcgtgcgtactggtcttgggtgctctaa 60
    |||
Db 73506 GACCGGGGGCGGTGGGTGACCCCGCTGCGTACTGCTTGTGGTGGCCCTTAA 73447

OY 61 tgtctgtctctaaagtcttgagggaagagcgggaggtctgtgcctgacactatg 120
    |||
Db 73446 TGTCTGTCTCTAAAGTCTTGAGGGGAAGACCGGGAGTCTGTGGCTTGAACCTATG 73387

OY 121 aaggaagagaacacacactctgcaggtgtgtgagaccacacgcctgaaagacgttg 180
    |||
Db 73386 AAGGAAGAGAACAACACTCGACGGTGTGACCAACCGCTGAAGACAGAGTGTG 73327

OY 181 ctggaagaagtcgccgaagaagt 202
    |||
Db 73326 CTGGAAGAAGTCCGCAAGAAGT 73305

RESULT 4
LOCUS 166494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
    source 1. 7218
    /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 2.8%; Score 58.8; DB 81; Length 7218;
Best Local Similarity 5.2%; Pred. No. 0.0012;
Matches 21; Conservative 224; Mismatches 161; Indels 0; Gaps 0;

OY 82 gaggggaaagacgcggaggtctgtgcctgacacatgaaagagaagatacaac 141
    |||
Db 1429 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1370

OY 142 ttgcaggtgtgagcaaccacgcctgaaacagctgtcgtgaaagaagtcgcgaagaag 201
    |||
Db 1369 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1310

OY 202 taggaagagaaacccgttatgagacacactgaataatgataagtagaagaaagca 261
    |||
Db 1309 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1250

OY 262 aattgtatctcaagagaagagtgaaataatcactgaagaaacttcaagatagacttc 321
    |||
Db 1249 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1190

OY 322 ttcttctgtctatgaagaaatccagtgtaactccttaacacattacttaagcaact 381
    |||
Db 1189 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1130

OY 382 agaagaagaaagacacttgaaagtaagtaataactatgacttaactgaagaaag 441
    |||
Db 1129 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1070

OY 442 agaataaagcttaacagaagatacaacatgaagcgcgtacatc 487
    |||
Db 1069 RRRATCGCAAGCTCCCTGACCTGACGCCAGCTCGAATTAATTC 1024

```

```

RESULT 5
LOCUS HS47C6R 318 bp DNA PRI 17-OCT-1995
DEFINITION H.sapiens Cpg Island DNA genomic MseI fragment, clone 47c6, reverse
    read cpq47c6.rta.
ACCESSION Z55522.1 GI:1021563
VERSION Cpg Island; genomic MseI fragment.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
    Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 318)
AUTHORS MacDonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE 2 (bases 1 to 318)
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEM-5zf(-)
    Clones are available from the UK MRC Human Genome Mapping Project
    Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
    http://www.hgmp.mrc.ac.uk/ for details
    or contact: biodelphgmp.mrc.ac.uk.
FEATURES
    source 1. 318
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /sex="male"
    /dev_stage="adult"
    /tissue_type="blood"
    /clone_id="CGI-1"
    /clone="47c6"
BASE COUNT 57 a 102 c 105 g 54 t
ORIGIN

Query Match 2.7%; Score 57; DB 77; Length 318;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaccggggggggttggttaccgctcgtgcgtactggtcttgggtgctctaa 57
    |||
Db 57 GACCGGGGGCGGTGGGTGACCCCGCTGCGTACTGCTTGTGGTGGCCCT 1

RESULT 6
ID E10125 standard; RNA; UNC; 2069 BP.
AC E10125;
SV E10125.1
XX
DE 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE DNA encoding an immunogenicity protein.
XX
KW JP 1995284392-A/1.
XX
OS unidentified
OC unclassified.
XX
RN [1]
RP 1-2069
RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;
RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND GENE
    RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOON";

```

```

RL Patent number JP1995284392-A/1, 31-OCT-1995.
RL DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.
XX
XX OS Leucocytozon caulleryi
CC PN JP 1995284392-A/1
CC PD 31-OCT-1995
CC PF 19-APR-1994 JP 1994080643
CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO
CC PC C12N15/09,A61K39/015,C12P21/02;
CC CC strandedness: Double;
CC CC topology: Linear;
CC CC key Location/Qualifiers
CC FH 1. .2069
CC FT source /organism="Leucocytozon caulleryi"
CC FT /strain="shizuoka"
CC FT misc_feature /note="EcORI adaptor"
CC FT misc_feature 2045. .2069
CC FT misc_feature /note="EcORI adaptor"
CC FT misc_feature 25. .2046
CC FT /product="immunogenicity protein"
XX
XX FH key Location/Qualifiers
XX FH source 1. .2069
XX FT /db_xref="taxon:32644"
XX FT /organism="unidentified"
XX
XX Sequence 2069 BP: 1216 A; 167 C; 443 G; 243 T; 0 other;

```

```

Query Match 2.3%; Score 48; DB 21; Length 2069;
Best Local Similarity 48.2%; Pred. No. 0.48;
Matches 164; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

OY 168 gaacacagcgttgcgtggaagtcgcgaagaagtatgtatgaaacccg-ttatgag 226
DB 1078 GAAGAAGAAAGAAAGTAACATGAAAGAAAGAAAGAAAGTAATACATGAAGAA 1137
OY 227 acacaactgaatgaatgaactagaagaagcaattgttatctcaaggagaagtg 286
DB 1138 AAGAAGAAAGTAATACATGAAGAAAGAAAGAAAGTAACATGAAGAAAGAA 1197
OY 287 gaaaaaatcattgaaaccttcgaagtagactatcttctatctggtctatgaagcgaatg 346
DB 1198 GAAGTAACATGAAGAAAGAAAGAAAGTAACATGAAGAAAGAAAGAAAGTA 1257
OY 347 ccagtggaatccttaaacattacttaacagctgaagaagaagaagcctcttga 406
DB 1238 ACACATGAAGAAAGAAAGAAAGTAACATGAAGAAAGAAAGTAACATGA 1317
OY 407 agtcaagtgaatactatgcacttaaacctgaagaagaagaagcttacagaagatc 466
DB 1318 GAAGAAGAAAGAAAGTAACATGAAGAAAGAAAGAAAGTAACATGAAGAA 1377
OY 467 aacaatgaagcgcgtatatacctagctgaatgctcagg 506
DB 1378 GAAAAAGTAACATGAAGAAAGAAAGTAACATG 1417

RESULT 7
ID E10126 standard; DNA; UNC; 3399 BP.
XX
XX AC E10126;
XX
XX SV E10126.1
XX
XX 08-OCT-1997 (Rel. 52, Created)
XX 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
XX DNA encoding an immunogenicity protein of Leucocytozon caulleryi fused to

```

```

DE maltose-binding protein.
XX
XX KM JP 1995284392-A/2.
XX
XX OS unidentified.
XX OC unclassified.
XX
XX [1]
XX RN 1-3399
XX RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;
XX RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZON IMMUNOGENIC PROTEIN AND GENE
XX RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZON";
XX RL Patent number JP1995284392-A/2, 31-OCT-1995.
XX DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.
XX
XX OS None
XX CC Artificial sequences.
XX CC PN JP 1995284392-A/2
XX CC PD 31-OCT-1995
XX CC PF 19-APR-1994 JP 1994080643
XX CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO
XX CC PC C12N15/09,A61K39/015,C12P21/02;
XX CC CC strandedness: Double;
XX CC CC topology: Linear;
XX CC FH key Location/Qualifiers
XX CC source 1. .3399
XX CC /organism="Artificial sequences"
XX CC FT CDS 1. .3399
XX CC /product="fusion protein of maltose-binding
XX CC protein and an
XX CC immunogenicity protein"
XX CC FT misc_feature 1. .1149
XX CC /note="maltose-binding protein"
XX CC FT misc_feature 1150. .1174
XX CC /note="EcORI adaptor"
XX CC FT misc_feature 1174. .3195
XX CC /note="immunogenicity protein"
XX CC FT misc_feature 3194. .3218
XX CC /note="EcORI adaptor"
XX CC FT misc_feature 3219. .3399
XX CC /note="sequence derived from pMAL-c vector"
XX
XX FH key Location/Qualifiers
XX FH source 1. .3399
XX FH /db_xref="taxon:32644"
XX FH /organism="unidentified"
XX
XX Sequence 3399 BP: 1577 A; 509 C; 797 G; 516 T; 0 other;

```

```

Query Match 2.3%; Score 48; DB 21; Length 3399;
Best Local Similarity 48.2%; Pred. No. 0.48;
Matches 164; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

OY 168 gaacacagcgttgcgtggaagtcgcgaagaagtatgtatgaaacccg-ttatgag 226
DB 2227 GAAGAAGAAAGAAAGTAACATGAAGAAAGAAAGAAAGTAATACATGAAGAA 2286
OY 227 acacaactgaatgaatgaactagaagaagcaattgttatctcaaggagaagtg 286
DB 2287 AAGAAGAAAGTAATACATGAAGAAAGAAAGAAAGTAACATGAAGAAAGAA 2346
OY 287 gaaaaaatcattgaaaccttcgaagtagactatcttctatctggtctatgaagcgaatg 346
DB 2347 GAAGTAACATGAAGAAAGAAAGAAAGTAACATGAAGAAAGAAAGTAACAT 2406
OY 347 ccagtggaatccttaaacattacttaacagctagaagaagaagaagcctcttga 406
DB 2407 ACACATGAAGAAAGAAAGTAACATGAAGAAAGAAAGTAACATGAAGAA 2466
OY 407 agtcaagtgaatactatgcacttaaacctgaagaagaagaagcttacagaagatc 466

```

Db 2467 GAAGAAAAAGAGAGATGACATGAGAGAAAAAGAGATGACATGAGAGAA 2526
 Oy 467 aacatgaacgcctacatagctgaatgctcag 506
 Db 2527 GAAAGATGACATGAGAGAGAAAAAGATGACATG 2566

RESULT 8
 AC004520/c LOCUS AC004520 132513 bp DNA PRI 03-FEB-2000
 DEFINITION Homo sapiens BAC clone CTB-119C2 from 7p15, complete sequence.
 AC004520
 VERSION AC004520.1 GI:3004572
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 132513)
 Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
 The sequence of Homo sapiens BAC clone CTB-119C2
 Unpublished
 2 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (01-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (03-FEB-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 1 (bases 1 to 132513)
 Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
 The sequence of Homo sapiens BAC clone CTB-119C2
 Unpublished
 2 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (01-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (03-FEB-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 1 (bases 1 to 132513)
 Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
 The sequence of Homo sapiens BAC clone CTB-119C2
 Unpublished
 2 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (01-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (03-FEB-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/CTB/CHR/, send
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 Clone CTB-119C2 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).
 VECTOR: pBelobAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The actual start of this clone is at base position 1 of CTB-119C2; actual end is at 132513 of CTB-119C2. The orientation of this clone is unknown.

This clone contains STS SWS2790 (NID:g1113585).

FEATURES
 Location/Qualifiers
 1..132513

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p15"
 /clone="CTB-119C2"
 /clone_1id="CTB-978SK-B"
 /complement(92..580)
 /note="match to EST AA53630 (NID:g2292107) vk57e11.r1"
 /complement(1168..1437)
 /note="match to EST H99041 (NID:g1123709) yx11h12.s1"
 /complement(1202..1437)
 /note="match to EST R66707 (NID:g839345) y133d08.s1"
 /complement(1272..1437)
 /note="match to EST AA036731 (NID:g1509988) zk30c06.s1"
 /complement(1293..1437)
 /note="match to EST AA745905 (NID:g2785891) ob02h02.s1"
 /complement(1305..1437)
 /note="match to EST AA701656 (NID:g2704821) z143f09.s1"
 /complement(1322..1437)
 /note="match to EST N27564 (NID:g1142045) yy01h07.s1"
 /complement(1349..1437)
 /note="match to EST AA678000 (NID:g2658522) z114f08.s1"
 /complement(1370..1437)
 /note="match to EST AA788753 (NID:g2848873) ah30g06.s1"
 1380..1437
 /note="match to EST (NID:g1434444)"
 /complement(1386..1437)
 /note="match to EST AA150090 (NID:g1721330) z142a06.s1"
 /complement(1393..1437)
 /note="match to EST AA701656 (NID:g2704821) z143f09.s1"
 /complement(1393..1437)
 /note="match to EST R66707 (NID:g839345) y133d08.s1"
 /complement(1393..1437)
 /note="match to EST N27564 (NID:g1142045) yy01h07.s1"
 /complement(1393..1437)
 /note="match to EST AA678000 (NID:g2658522) z114f08.s1"
 /complement(1393..1437)
 /note="match to EST H99041 (NID:g1123709) yx11h12.s1"
 /complement(1393..1437)
 /note="match to EST AA150090 (NID:g1721330) z142a06.s1"
 /complement(1393..1437)
 /note="match to EST AA036731 (NID:g1509988) zk30c06.s1"
 /complement(1393..1437)
 /note="match to EST AA745905 (NID:g2785891) ob02h02.s1"
 1963..2262
 /note="match to EST (NID:g1434444)"
 /complement(1963..2301)
 /note="match to EST AA788753 (NID:g2848873) ah30g06.s1"
 3282..3373
 /rpt_family="MER2-type"
 4620..4819
 /rpt_family="MER2-type"
 4925..5022
 /rpt_family="Alu"
 6135..6571
 /note="match to EST W76339 (NID:g1386789) zd60f11.s1"
 6149..6531
 /note="match to EST H21401 (NID:g890096) y132g06.s1"
 /complement(6641..39925)
 /gene="WUSC:H_RG119C02.1"
 /complement(join(6641..7891,8640..8723,14302..14481,39356..39925))
 /gene="WUSC:H_RG119C02.1"
 /note="similar to NFE2-related transcription factors;

```

similar to I48694 (PID:g2137676); H_RG119C02.1"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC09039.1"
/db_xref="GI:3004573"
/translation="MKHLKRWMSAGGGLLHLTLISLAGLRVDLDLYLLPPTLLD
ELPLFGPRASSAYALSPSPSASGGMRAGHLHKGKRELDPAAPPECOLUREVALICVPR
VPRISVDAMLVHVASGASDEAHGILLGAASSTGAGASVDGSOAVQGGGDPRAA
RSGPLDAGEEERAPAPTAQVPDAGCASAEEGVLRKEHVAVHSSOHEENERVSAQ
KENSILQDDDDENKIAERPDWEAKETESRNRHNGTDFSLDLQLSSQSPENS
LEGISLIDPLPESISDGMNSAHYVNSOAIQDVNHEAILLCPNTPFRDPTAR
TSOQEPPLQINSHTNPEQPLPGNTLGFSLSPVNHMNTLSODLLDIDINDEI
NLMGATEDNPPPIVSOLFDEPDSDGSLSDSNNNSVITSNSHSDCDGAGICY
TDHSSSHHDEGAVGTYPEPSKLCILDQSDSDHGDUTFQVHFNHTYHLQPTAPE
STSEPPWPGKSKQKIRSRLEDYEDNRNLSDERAKALHLPESVDIYVPMDFSM
SRVYLDQVSLRIDIRRGKNVAQNRKRRLDIILLEDVQNLQAKKETLRQO
AOCNAINMKOKLHDLYHDIFSRRLRDDGGRVNNHVALQCTHDGSLIYKELVAS
GHKKTOKGRK"
complement(7777..7893)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA135857 (NID:g1697161) z025h07.r1"
complement(7781..7897)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA149705 (NID:g1720647) zn99f04.r1"
complement(7798..7897)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA099770 (NID:g1646643) z181a01.r1"
8305..8453
/rpt_family="L1"
complement(8598..8725)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA132584 (NID:g1694136) z020c03.r1"
complement(8602..8862)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST I44351 (NID:g1048879)"
8865..9158
/rpt_family="Alu"
9338..9614
/rpt_family="Alu"
9792..10082
/rpt_family="Alu"
11148..11429
/rpt_family="Alu"
12696..12873
/rpt_family="Alu"
12908..13196
/rpt_family="Alu"
complement(14301..14483)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA149705 (NID:g1720647) zn99f04.r1"
complement(14302..14483)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA099770 (NID:g1646643) z181a01.r1"
complement(14302..14443)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA135857 (NID:g1697161) z025h07.r1"
complement(14302..14483)
/gene="WUGSC:H_RG119C02.1"
/feature="match to EST AA132584 (NID:g1694136) z020c03.r1"
14796..16547
/rpt_family="SVA"
16630..16927
/rpt_family="Alu"
17052..17359

```

```

Query Match 2.1%; Score 45; DB 8; Length 132513;
Best Local Similarity 52.4%; Pred. No. 2.5;
Matches 99; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

```

QY 1004 tgccttcagagagctcctacagcaacctggatattcttactatgatctatttgcct 1063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126022 TGACTAGTGTGAGCTCTCTCCGTCACCTCAGTAATGTCTCAGATCAGTGGCC 125963

```

```

QY 1064 gctctgaactcttaaggaatgcaatgagcttgaaataatgcttcgtgtcagaattcg 1123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125962 ATGTAGATGACTTAAAGAAAAGCTCTTGGACTGAATTAACATTAATAAACCTATTGG 125903

QY 1124 agaaacaacaagaatctctctctcattcaatcaagaagaacaagaagaaga 1183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125902 TGGAGCCCAATTAATAAAGCATATTTCATTGGTGCAAATCTTCAGAAAGGACATTA 125843

QY 1184 tcaaatatt 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125842 TTCATTATT 125834

RESULT 9
AC022764 188563 bp DNA HTG 16-MAR-2000
LOCUS Homo sapiens chromosome 17 clone RP11-310M15 map 17, WORKING DRAFT
DEFINITION
ACCESSION AC022764.3 GI:7249221
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 188563)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-310M15
Unpublished
2 (bases 1 to 188563)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckel,Y., Bida,F.,
Boguslavsky,L., Bouckhagalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardina,S., Grant,G., Hagos,B., Hatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,B., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olyar,T.M., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rotman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,D., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced g1:6983429.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Project -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: 15848
Center clone name: 310.M.15
----- Summary Statistics -----
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 164966 bases at least Q40
Consensus quality: 176314 bases at least Q30
Consensus quality: 180940 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 184663; sum-of-ctnigs
Quality coverage: 2.7 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-ctnigs

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1196: contig of 1196 bp in length
* 1197 1296: gap of 100 bp
* 1297 2422: contig of 1126 bp in length
* 2423 2522: gap of 100 bp
* 2523 3938: contig of 1316 bp in length
* 3939 5333: contig of 1395 bp in length
* 5334 5433: gap of 100 bp
* 5434 6892: contig of 1459 bp in length
* 6893 6992: gap of 100 bp
* 6993 9051: contig of 1959 bp in length
* 9052 10497: contig of 1446 bp in length
* 10498 10597: gap of 100 bp
* 10598 12563: contig of 1966 bp in length
* 12564 12663: gap of 100 bp
* 12664 14771: contig of 2108 bp in length
* 14772 14871: gap of 100 bp
* 14872 16733: contig of 1862 bp in length
* 16734 16833: gap of 100 bp
* 16834 19429: contig of 2596 bp in length
* 19430 19529: gap of 100 bp
* 19530 21429: contig of 1900 bp in length
* 21430 21529: gap of 100 bp
* 21530 24676: contig of 3147 bp in length
* 24677 24776: gap of 100 bp
* 24777 27200: contig of 2424 bp in length
* 27201 27300: gap of 100 bp
* 27301 29853: contig of 2553 bp in length
* 29854 29953: gap of 100 bp
* 29954 32573: contig of 2620 bp in length
* 32574 32673: gap of 100 bp
* 32674 34834: contig of 2161 bp in length
* 34835 34934: gap of 100 bp
* 34935 38184: contig of 3250 bp in length
* 38185 38284: gap of 100 bp
* 38285 41429: contig of 3145 bp in length
* 41430 41529: gap of 100 bp
* 41530 43659: contig of 2130 bp in length
* 43660 43759: gap of 100 bp
* 43760 46302: contig of 2543 bp in length
* 46303 46402: gap of 100 bp
* 46403 49933: contig of 3531 bp in length
* 49934 50033: gap of 100 bp
* 50034 54240: contig of 4207 bp in length
* 54241 54340: gap of 100 bp
* 54341 57127: contig of 2787 bp in length
* 57128 57227: gap of 100 bp
* 57228 61476: contig of 4249 bp in length
* 61477 61576: gap of 100 bp
* 61577 65550: contig of 3974 bp in length
* 65551 65650: gap of 100 bp
* 65651 70580: contig of 4930 bp in length
* 70581 70680: gap of 100 bp
* 70681 76028: contig of 5348 bp in length
* 76029 76128: gap of 100 bp
* 76129 83825: contig of 7697 bp in length
* 83826 83925: gap of 100 bp
* 83926 90939: contig of 7014 bp in length
* 90940 91039: gap of 100 bp
* 91040 96426: contig of 5387 bp in length
* 96427 96526: gap of 100 bp
* 96527 101964: contig of 5438 bp in length
* 101965 102064: gap of 100 bp

* 102065 108545: contig of 6481 bp in length
* 108546 108645: gap of 100 bp
* 108646 114408: contig of 5843 bp in length
* 114409 114488: gap of 100 bp
* 114489 122283: contig of 7695 bp in length
* 122284 122383: gap of 100 bp
* 122384 131395: contig of 9012 bp in length
* 131396 131495: gap of 100 bp
* 131496 142144: contig of 10649 bp in length
* 142145 142244: gap of 100 bp
* 142245 154513: contig of 12269 bp in length
* 154514 154613: gap of 100 bp
* 154614 169786: contig of 15173 bp in length
* 169787 169886: gap of 100 bp
* 169887 188563: contig of 18677 bp in length.
Location/Qualifiers
1. 188563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-310M15"
/clone_11b="RP11 Human Male BAC"
1. 1196
/note="assembly_fragment"
1297. 2422
/note="assembly_fragment"
2523. 3638
/note="assembly_fragment"
3939. 5333
/note="assembly_fragment"
5434. 6892
/note="assembly_fragment"
6993. 8951
/note="assembly_fragment"
9052. 10497
/note="assembly_fragment"
10598. 12563
/note="assembly_fragment"
12664. 14771
/note="assembly_fragment"
14872. 16733
/note="assembly_fragment"
16834. 19429
/note="assembly_fragment"
19530. 21429
/note="assembly_fragment"
21530. 24676
/note="assembly_fragment"
24777. 27200
/note="assembly_fragment"
27301. 29853
/note="assembly_fragment"
29954. 32573
/note="assembly_fragment"
32674. 34834
/note="assembly_fragment"
34935. 38184
/note="assembly_fragment"
38285. 41429
/note="assembly_fragment"
41530. 43659
/note="assembly_fragment"
43760. 46302
/note="assembly_fragment"
clone_end:17
vector_side:left
46403. 49933
/note="assembly_fragment"

Query Match 2.1% Score 45; DB 50; Length 188563;
Best Local Similarity 48.3%; Pred. No. 2.5; Mismatches 135; Indels 0; Gaps 0;
Matches 126; Conservative 0;

```

OY 238 attaatgatgaactagaagaacaaattgttattctcaaggagaagtgaaaaatcca 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116423 ATTGAATGCTGGAATAAATTATACAGAAAAATCAATACGGCAAAAAATACAGAAAAAGAG 116482
OY 298 tgaacactcctgagatgatatcttctatcggtctatgaacgaatgcgaagcgatc 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116483 ATAGTAATATACAGAGAAATCAGGTAGAGAGCTGTGAATATGACAAATGCAAGAGAAAA 116542
OY 358 cttaacacactactactaaagctagaagaagaagaagactcttgaagcaatgaa 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116543 GGAATAATGATTAATAAACACACATGAAAGAGAAATTTAACTTTTGAGATTAATGAT 116602
OY 418 atactatgcactaaactgaacagaagaagcttaccagaagatcaacatgaacg 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116603 TATCTACCCGACACATCCAAACATATCTGACATCTGTGATATACAGAAATGAGATA 116662
OY 478 ccgtactactactagctgaat 498
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116663 TTCAGGTACTCTGCACGAAAT 116683

RESULT 10
AC034122 201010 bp DNA HTG 26-MAY-2000
LOCUS Mus musculus clone RP23-169C19, WORKING DRAFT SEQUENCE, 30
DEFINITION unoriented pieces.
AC034122
VERSION AC034122.3 GI:8077093
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201010)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-169C19
Unpublished
2 (bases 1 to 201010)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teste,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7770516.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research.
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: 18058
Center clone name: 109_C19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 184191 bases at least Q40
Consensus quality: 182258 bases at least Q30
Consensus quality: 195732 bases at least Q20
Insert size: 208000; agarose-fp
Insert size: 198110; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1111: contig of 1111 bp in length
* 1112 1211: gap of 100 bp
* 1212 2404: contig of 1193 bp in length
* 2405 2504: gap of 100 bp
* 2505 3976: contig of 1472 bp in length
* 3977 4076: gap of 100 bp
* 4077 5334: contig of 1258 bp in length
* 5335 5434: gap of 100 bp
* 5435 6696: contig of 1262 bp in length
* 6697 6796: gap of 100 bp
* 6797 9202: contig of 2406 bp in length
* 9203 9302: gap of 100 bp
* 9303 10287: contig of 985 bp in length
* 10288 10387: gap of 100 bp
* 10388 12441: contig of 2054 bp in length
* 12442 12541: gap of 100 bp
* 12542 14480: contig of 1939 bp in length
* 14481 14580: gap of 100 bp
* 14581 16881: contig of 2301 bp in length
* 16882 16981: gap of 100 bp
* 16982 18958: contig of 1977 bp in length
* 18959 19058: gap of 100 bp
* 19059 21985: contig of 2927 bp in length
* 21986 22085: gap of 100 bp
* 22086 24348: contig of 2263 bp in length
* 24349 24448: gap of 100 bp
* 24449 28133: contig of 3685 bp in length
* 28134 28233: gap of 100 bp
* 28234 32277: contig of 4044 bp in length
* 32278 32377: gap of 100 bp
* 32378 38555: contig of 6178 bp in length
* 38556 38655: gap of 100 bp
* 38656 44903: contig of 6248 bp in length
* 44904 45003: gap of 100 bp
* 45004 52544: contig of 7541 bp in length
* 52545 52644: gap of 100 bp
* 52645 61054: contig of 8410 bp in length
* 61055 61154: gap of 100 bp
* 61155 68866: contig of 7732 bp in length
* 68867 68966: gap of 100 bp
* 68967 77233: contig of 8247 bp in length
* 77234 77333: gap of 100 bp
* 77334 86303: contig of 8970 bp in length
* 86304 86403: gap of 100 bp
* 86404 96573: contig of 10170 bp in length
* 96574 96673: gap of 100 bp
* 96674 107775: contig of 11102 bp in length
* 107776 107875: gap of 100 bp
* 107876 119036: contig of 11161 bp in length
* 119037 119136: gap of 100 bp
* 119137 133728: contig of 14592 bp in length

```

```

* 133729 133828: gap of 100 bp
* 133829 145758: contig of 11930 bp in length
* 145759 145858: gap of 100 bp
* 145859 161359: contig of 15501 bp in length
* 161360 161459: gap of 100 bp
* 161460 178068: contig of 16609 bp in length
* 178069 178168: gap of 100 bp
* 178169 201010: contig of 22842 bp in length.
FEATURES
    Location/Qualifiers
        1..201010
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="RP23-169C19"
        /clone_lib="RP23-169C19"
        1..1111
        /note="assembly-fragment"
misc_feature 1212..2404
    /note="assembly-fragment"
misc_feature 2505..3976
    /note="assembly-fragment"
misc_feature 4077..5334
    /note="assembly-fragment"
misc_feature 5435..6696
    /note="assembly-fragment"
misc_feature 6797..9202
    /note="assembly-fragment"
misc_feature 9303..10287
    /note="assembly-fragment"
misc_feature clone_end:T7
    vector_side:left"
    10388..12441
    /note="assembly-fragment"
misc_feature 12542..14480
    /note="assembly-fragment"
misc_feature 14581..16881
    /note="assembly-fragment"
misc_feature 16982..18958
    /note="assembly-fragment"
misc_feature 19059..21985
    /note="assembly-fragment"
misc_feature 22086..24348
    /note="assembly-fragment"
misc_feature 24449..28133
    /note="assembly-fragment"
misc_feature 28234..32277
    /note="assembly-fragment"
misc_feature 32378..38555
    /note="assembly-fragment"
misc_feature 38656..44903
    /note="assembly-fragment"
misc_feature 45004..52544
    /note="assembly-fragment"
misc_feature 52645..61054
    /note="assembly-fragment"
misc_feature 61155..68886
    /note="assembly-fragment"
misc_feature clone_end:SP6
    vector_side:left"
    68987..77233
    /note="assembly-fragment"
misc_feature 77334..86303
    /note="assembly-fragment"
misc_feature 86404..96573
    /note="assembly-fragment"
misc_feature 96674..107775
    /note="assembly-fragment"
misc_feature 107876..119036
    /note="assembly-fragment"
misc_feature 119137..133728
    /note="assembly-fragment"
misc_feature 133829..145758
    /note="assembly-fragment"
misc_feature 145859..161359
    /note="assembly-fragment"

```

```

Query Match      2.1%; Score 44; DB 54; Length 201010;
Best Local Similarity 50.5%; Pred. No. 4.4;
Matches 107; Conservativity 0; Mismatches 105; Indels 0; Gaps 0;

misc_feature      /note="assembly-fragment"
                  161460..178068
                  /note="assembly-fragment"
misc_feature      /note="assembly-fragment"
                  178169..201010
                  /note="assembly-fragment"

```

```

RESULT 11
AC026784
LOCUS       AC026784      132491 bp      DNA
DEFINITION Homo sapiens chromosome 5 clone CTD-2138014, WORKING DRAFT
ACCESSION   AC026784
VERSION     AC026784.3  GI:9964791
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 132491)
AUTHORS    DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 132491)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL     Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
COMMENT     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Sep 2, 2000 this sequence version replaced gi:7711994.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 680891
            Center clone name: CTRB-H1_2138014
            -----
            Summary Statistics
            Consensus quality: 124042 bases at least Q40
            Consensus quality: 129692 bases at least Q30
            Consensus quality: 130729 bases at least Q20
            Estimated insert size: 132010; agarose-fp estimation
            Estimated insert size: 131991; sum-of-ctdigs estimation
            Quality coverage: 6.48 in Q20 bases; agarose-fp estimation
            Quality coverage: 6.48 in Q20 bases; sum-of-ctdigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced

```

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 162162)	Bairren,B., Linton,L., Nusbaum,C. and Lander,E.	Homio sapiens, clone RP11-10119	Unpublished	
2 (bases 1 to 162162)	Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barrna,N., Beckerly,R., Boguslavsky,L., Bouknighter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gaidyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kahn,L., Karkats,A., Klein,J., Lechnocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., Mcwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,D., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Sudriaman,A., Talamas,J., Testae,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.	Submitted (27-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	Submitted	
3	On Mar 12, 2000 this sequence version replaced g1:6649424.	All repeats were identified using RepeatMasker:		
4	Smit, A.F.A. & Green, P. (1996-1997)	http://ftp.genome.washington.edu/BM/RepeatMasker.html		
5	Center: Whitehead Institute/ MIT Center for Genome Research			
6	Center code: WtBR			
7	Web site: http://www-seq.wi.mit.edu			
8	Contact: sequence_submissions@genome.wi.mit.edu			
9	Project Information			
10	Center project name: L3075			
11	Center clone name: 10.I.19			

```

Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 157087 bases at least Q40
Consensus quality: 159567 bases at least Q30
Consensus quality: 160677 bases at least Q20
Insert size: 170000; agarose-fp
Quality size: 161462; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1          4904: contig of 4904 bp in length
* *          4905 5004: gap of          100 bp
* *          5005 12177: contig of 7173 bp in length
* *          12178 12277: gap of          100 bp
* *          12278 19252: contig of 6975 bp in length

```

```

* 30546 30653: gap of 100 bp
* 30646 43678: contig of 13033 bp in length
* 43679 43778: gap of 100 bp
* 43779 71005: contig of 27221 bp in length
* 71006 71105: gap of 100 bp
* 71106 116667: contig of 45562 bp in length
* 116668 116767: gap of 100 bp
* 116768 162162: contig of 45395 bp in length.
* Location/Qualifiers
  1..162162
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
FEATURES
SOURCE
```


clone

1	4200:	contig of 4200 bp in length
4201	4300:	gap of unknown length
4301	8950:	contig of 4651 bp in length
8952	9051:	gap of unknown length
9052	14534:	contig of 5463 bp in length
14535	14634:	gap of unknown length
14635	23344:	contig of 8710 bp in length
23345	23444:	gap of unknown length
23445	33995:	contig of 10551 bp in length
33996	34095:	gap of unknown length
34096	44270:	contig of 10175 bp in length
44271	44370:	gap of unknown length
44371	62317:	contig of 17947 bp in length
62318	62417:	gap of unknown length
62418	90640:	contig of 28223 bp in length
90641	90740:	gap of unknown length
90741	119137:	contig of 28397 bp in length
119138	119237:	gap of unknown length
119238	164615:	contig of 45582 bp in length
164620	164715:	gap of unknown length
164720	167612:	contig of 2853 bp in length
167613	167712:	gap of unknown length
167713	171189:	contig of 4007 bp in length
171189	gap of	unknown length
171820	contig of	2238 bp in length

UNES	location/Qualifiers
source	1..174057 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /clone="RP11.645N11" 1..4200
misc_feature	/note="assembly_name:Contig104301..8951
misc_feature	/note="assembly_name:Contig119052..14534
misc_feature	/note="assembly_name:Contig12

Query Match	2.1%	Score	43.8	DB	58	Length	174057
Best Local Similarity	53.9%	Pred	NO.4.9				
Matches	90	Conservative	0	Mismatches	77	Indels	0
						Gaps	0

RESULT 15
PCU43145

LOCUS	PCU43145	9965 bp	DNA	INV	08-OCT-1998
DEFINITION	Plasmidium chabaudi repeat organelle protein gene, complete cds				
ACCESSION	U43145				
VERSION	U43145.1	GI:1151157			
KEYWORDS					
SOURCE	Plasmidium chabaudi.				
ORGANISM	Plasmidium chabaudi.				

AUTHORS	JOURNAL	REFERENCE	AUTHORS	JOURNAL
Werner, E. B., Holder, A. A., Asroodi, A., and Taylor, W. R.				
A novel 11-residue coiled-coil motif predicts a histidine zipper				
Protein Pept. Lett. 3 (2), 139-146 (1996)				
2 (bases 1 to 9985)				
Werner, E. B., Taylor, W. R., and Holder, A. A.				
A plasmidom chabaudi protein contains a repetitive region with a				
predicted spectrin-like structure				
Mol. Biochem. Parasitol. 94 (2), 185-196 (1999)				

MEDLINE
98418/05
3 (bases 1 to 9985)
REFERENCE

AUTHORS Werner, E. B.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1995) Ekkehard B. Werner, Parasitology, National

```
FEATURES      Location/Qualifiers
source        1..9985
              /organism="Plasmodium chabaudi"
              /strain="96V"
```

```

/note="ROPE"
/codon_start=1
/evidence=experimental
/product="repeat Organellar protein"
/protein_id="AAC63403.1"
/db_xref="GI:1151158"

```

repeat_region	4294	.	5385
	/rpt_type=tandem		
BASE COUNT	4487	a	766 c 1644 g 3088 t
ORIGIN			

Query Match	2.0%;	Score 43.6;	DB 56;	Length 9985;
Best Local Similarity	47.7%;	Pred. NO. 5.5;		
Matches 127;	Conservative 0;	Mismatches 139;	Indels 0;	Gaps 0;

QY 185 aagagtcgcgaagaagtatgtaatgtaaaaccgttatgagacaacttgaaat 244
||| || | ||| ||| ||| ||| ||| ||| ||| |||
Db 6451 AATAATTAAATTAACCACTGAGATGAAGACACGGTGATGATTAATTTGCAAAAAAAT 651

[illegible]

OY. 305 tcttcagatagactatacttcatcgtgctctatgaacgaatgccagtggtaattccctaacc 364
| | | | |
Dh 6571 TATCCCTGAGACAAATTAAAAAATAACCAATGCATCAGAAAACCTAAAAATTAACTTAAC 6630

QY 365 acattacttaacagctagaagaagaagacgtcttgaagtcaagtgaatactat 424

QY 425 gcacttaactggaacaagaatcaaa 450
||| ||| ||||| |||

Db 6691 AACTGAGAGATATTAAAGAAAAA 6716

Search completed: January 19, 2001, 02:57:41
Job time: 13829 sec

Tue Jan 23 10:54:39


```

2   TTAAGTGAACACTGACATGAGAAAAATGACTAATGTTGTTAAATACTG 51
70  pThrThCysProSerAsnThrValThyGluAsnAlaThrGlyProG 87
52  GACAAACATGTCCTCAACACTGTTAGACTGAAACGCAACTGGGCTG 101
87  IugluLeuGlyLeuProLeuGlnArgSerTySerGluHisLeuGlyTyr 103
102  AAGAACTTGATTCCTCCCTGACAGAGTCTTACACGAAACACCTGGATAT 151
104  PheProThrAspLeuPheAlaCysSerGluSerLeuArgAsnGlyAsnG 120
152  TTCTCTACTGATCTATTGCTCTGCTGAAATCTTTAAGAAATGCAATGG 201
120  yLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsnLysLysL 137
202  GCTTGAATTAATGCTTGTGTCAGAGTTGCAAAAAAAGAAAGATCT 251
137  eLeuLeuHisSerSerLysGluLysLeuArgGluArgGlyLeuTyr 153
252  CTCTCTTCATTCACAGCAAGAAAACTAAGAAAGCAAGAAATCAATAT 301
154  CysCysGluGlnLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
302  TGCTGTAGCAGCTGCTACTCTCTGCTCCGATGTAAAGGAGGAGAAAGA 351
170  snAspAlaAlaSerValLeuGluAlaThrValAspTyrValLysTyr 186
352  ATGATGGGGCTTCAGTCTTGAGGCAACAGTATTTATGTGAATATA 401
186  leArgGluLysLysSer 191
402  TCCCGGAGAAATCTCT 419

seq_name: gb_est7:AA417643

seq_documentation_block:
LOCUS   AA417643      490 bp      mRNA      EST      02-MAR-1998
DEFINITION   zv04a08.r1 Soares.NHMPU.S1 Homo sapiens cDNA clone IMAGE:752630
5', mRNA sequence.
ACCESSION   AA417643
VERSION     AA417643.1 GI:2079462
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 490)
AUTHORS    Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theising,B.,
White,X., Wylie,T., Waterston,R. and Wilson,R.
WashU-MC1 human EST Project
JOURNAL    Unpublished (1997)
COMMENT     Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
314 286 1800
314 286 1810
est@wustl.edu
is available royalty-free through LNL; contact the
LNL (info@lml.lnl.gov) for further information.
Std Error: 0.00
313 rev2 ET from Amersham
ence stop: 465.
Qualifiers
" Homo sapiens "
" 5975465 "
" 0606 "
" NHMPU.S1 "
human melanocyte, fetal heart, and

```

```

pregnant uterus"
/lab_host="DB10B"
/note="Organ: mixed (see below); Vector: pF773D-Pac
(pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBH, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      158 a      114 c      97 g      121 t

alignment_scores:
Quality: 386.00      Length: 73
Ratio: 5.288      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 98.630

alignment_block:
US-09-389-000-2 x AA417643 ..
Align seg 1/1 to: AA417643 from: 1 to: 490

333 PheserleuGlyThrAlaLeuLeuGlyTyrPalaArgAlaLeuHisL 349
|||||
3 TTTCTCCTCGGCACTGCATCTTGCGTGGGTGAGCACTGCATACAT 52
349 eProthValCysAsnSerPheGlyArgIleLysSerThrCysLeuLysP 366
|||||
53 CCCAAGCTGTGCAACAGTTTGGGCTATTAAAGCACATGTTGAAAT 102
366 heThrLeuSerThrThrTyrTyrPalaGlnPheAspAsnLeuGlyVal 382
|||||
103 TCACACTCTCAACACCACTACTGCGGCAGCTTGACATCAGCAAAAAAGT 152
383 GlnGlnArgMetIleLeuLysAlaProProlAspLeuLeuSerLysG 399
|||||
153 GAACAAAGAAATGATTTGAAAGCTCCACCCCAAGACCTATATCAAAAGA 202
399 uLeuAlaTrrPheGlyPhe 405
|||||
203 GTTGGCATGCTTTGGCTTC 221

seq_name: gb_est51:AW961069

seq_documentation_block:
LOCUS   AW961069      617 bp      mRNA      EST      01-JUN-2000
DEFINITION   EST373036 MAGE resequences; MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW961069
VERSION     AW961069.1 GI:8150648
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 617)
AUTHORS    Hegde,P., Qi,R., Abernathy,K., Dharp,S., Gaspar,D., Gay,C., Holt
J.E., Saged,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL    Unpublished (2000)
COMMENT     Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: john@tifg.org
Plate: 152

```

The Jan 23 10:54:46 2001

Om of. Hs. 10:54:46 2001

Human melanocyte, fetal heart, and

Seq primer: Forward.

FEATURES Location/Qualifiers
Source 1..617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE Resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 211 a 116 c 127 g 163 t
ORIGIN

alignment_scores:

Quality: 291.00 Length: 59
Ratio: 4.932 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:

US-09-389-000-2 x AW961069 ..

Align seg 1/1 to: AW961069 from: 1 to: 617

```

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGluAl 17
|||||
439 ATGCTCTCTTGAAGTGGCTTCTTCACTAAGTCCGAGAGGCTGGAGC 488
|||||
17 aAlLeuLeuLeuLeuArpPheGlyLysLysLysAsnThrHisSerLeuPhe 34
|||||
489 CACCAAGTAAATAGATTGGCAAAAAGAAAATACACATTCACTGTTG 538
|||||
34 aAlPheLeuLeuProGluAsnPhelLysGlyLysIleSerGlyHisGlyMet 50
|||||
539 TTTTATTAATCCCTGAAATTTTAAAGTTGTATTCAGGGCATGAGATG 588
|||||
51 AspLeuAlaLeuThrGluProLeuThr 59
|||||
589 GATATGCTTTAACTGAACCACTGACA 615

```

seq_name: gb_est74:BE665112

seq_documentation_block:

LOCUS BE665112 503 bp mRNA EST 08-SEP-2000
DEFINITION 153097 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE665112
VERSION BE665112.1 GI:10025321

KEYWORDS

EST.

SOURCE

cow.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 503)

REFERENCE

AUTHORS

Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.M. and Keeler,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

TITLE

JOURNAL

Unpublished (2000)
Contact: Smith TP
USDA, ARS US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

COMMENT

Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -m1nscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGAACAAGTATGACACAT
BACKWARD: GTTCCAGTCAGCAGC
Plate: 65 row: G column: 20
Seq primer: ATTTAGGTACACATATAG.
Location/Qualifiers
1..503

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 4BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."

BASE COUNT 159 a 108 c 101 g 135 t
ORIGIN

alignment_scores:

Quality: 258.50 Length: 90
Ratio: 3.541 Gaps: 3
Percent Similarity: 81.111 Percent Identity: 63.333

alignment_block:

US-09-389-000-2 x BE665112 ..

Align seg 1/1 to: BE665112 from: 1 to: 503

```

318 AsnIleSerIleHisLeuProSerAlaMetProProValSer SerPheS 334
|||||
3 AACAATTCAGTGGCTTCCACACAGCTGATCCAAAGTGTCAGATTCT 52
|||||
334 erLeuGlyThrAlaLeuLeuGlyTTPAlaArgArgAlaLeuHisIlePro 350
|||||
53 TCCTCAACATGCAATTCGTGTGGCCGAGTGTGCACAGCACACCACCA 102
|||||
351 ThrValCysAsnSerPheGlyArgIleLysSerThrCysLeuLysPheTh 367
|||||
103 ACTGCTGCACAGATTGGCGCTATTAATACACATATTGGAATGAC 152
|||||
367 rIleu.SerThrThrTyrrTPAlaGlnPheAspAsnLeuGlyLysValGlu 383
|||||
153 TCTAGACC...ACCTATGCGGTGCGCTGACATACAGGAAGAAATGAA 199
|||||
384 GlnArgMetIleLeuLysAlaProProLysAspLeuIleSerGlyGlu 400
|||||
200 CAAGAATGATCTTGAAGCTACACCCAGAGACTGATATCAAAAGGTT 249
|||||
400 uAlaTrpPheGlyPhe 405
|||||
250 TGCATTGTATGCGCTC 265

```

seq_name: gb_est51:AW961078

seq_documentation_block:

LOCUS AW961078 624 bp mRNA EST 01-JUN-2000
DEFINITION EST373045 MAGE Resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW961078
VERSION AW961078.1 GI:8150657

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 624)

REFERENCE

AUTHORS

Hedge,P., Qi,R., Abernathy,K., Dharp,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,Y., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

TITLE

Unpublished (2000)

JOURNAL

Contact: John Quackenbush

COMMENT

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@igr.org
Plate: 152
Seq primer: forward.

FEATURES
source
Location/Qualifiers
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGF"
/note="Vector: pBluescriptSKm"

BASE COUNT 209 a 117 c 128 g 170 t
ORIGIN

Alignment_scores:
Quality: 181.00 Length: 65
Ratio: 3.175 Gaps: 5
Percent Similarity: 87.692 Percent Identity: 78.462

Alignment_block:
US-09-389-000-2 x AW961078 ..

Align seg 1/1 to: AW961078 from: 1 to: 624

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGln 17
|||||
439 ATGGTCTCTTGAAGCTGCTTTTACCACTGCGAGAGCTGGAAC 488
17 alleLysLeuLeuArg.PheGlyLysLysAsnThr.HisSerLeu 33
|||||
489 CATCAAGTAAATTAGAAATTTGCCAAGAAATTCACAAATTCCTCTT 538
33 eValPheLeuLeuProGluAsnPheLysGlyCysIleSer.GlyHisGly 49
|||||
539 TGTTTTAAATCCCTGAAATTTTAAAGGTGTTTTCACAGGGCCTGA 588
50 MetAspLeuLeuThr.GluProLeuThrMet 60
|||||
589 AATGATTTGTTTAACTTGAACCCCTGACCAATG 624

seq_name: gb_gss3: A0196491

seq_documentation_block:
LOCUS A0196491 452 bp DNA GSS 16-SEP-1998
DEFINITION CIT-HSP-2383N20.TF CIT-HSP Homo sapiens genomic clone 2383N20, DNA
sequence.
ACCESSION A0196491
VERSION A0196491.1 GI:3603853
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 452)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,U.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other-GSS: CIT-HSP-2383N20.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hunggen/Bac_end_search/Bac_end_search.html.
Seq primer: M13-21
Class: BAC ends

FEATURES
source
Location/Qualifiers
1..452
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
Location/Qualifiers
1..624
/clone_lib="2383N20"
/db_xref="taxon:9606"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 128 a 88 c 88 g 148 t
ORIGIN

Alignment_scores:
Quality: 159.00 Length: 59
Ratio: 3.532 Gaps: 1
Percent Similarity: 76.271 Percent Identity: 57.627

Alignment_block:
US-09-389-000-2 x A0196491 ..

Align seg 1/1 to: A0196491 from: 1 to: 452

198 lIerhGluAlaLeuGlnSerAsnMetArgPheCysLysLysGlnGln 214
|||||
191 ATTACAGAGACACTTCACAGACATGAGGTTTGTAGAAACAAAC 240
214 rProLeuLeuSerLeuProGlyThrValMetAlaGlnArg..... 228
|||||
241 ACCCATTAAGCTGTCTCTCCACGACTGTATGCGACAGCGGTATGATA 290
229GluAsnSerValMetSerThrTyrSerProGluArgGly 241
|||||
291 AGGGCTTTGGCAATCTGAAGTCTGCTAGTATTTACTCTCCAGTTGAT 340
242 LeuGlnPheLeuThrAsnThrCysTrp 250
|||||
341 GCTAATATGTACACAAACTGCTAT 367

seq_name: gb_gss24: B92720

seq_documentation_block:
LOCUS B92720 354 bp DNA GSS 25-JUN-1998
DEFINITION CIT-HSP-2170M15.TF CIT-HSP Homo sapiens genomic clone 2170M15, DNA
sequence.
ACCESSION B92720
VERSION B92720.1 GI:2975057
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 354)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,U.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hunggen/Bac_end_search/Bac_end_search.html
Seq primer: M13-21;
Class: BAC ends

FEATURES
source
Location/Qualifiers
1..354
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

/clone="2170M15"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBelobAC11. Site_1: HindIII; Site_2:
HindIII"
BASE COUNT      93 a      70 c      67 g      124 t
ORIGIN

```

```

alignment_scores:
  Quality: 146.00      Length: 30
  Ratio: 4.867         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.667

```

```

alignment block:
US-09-389-000-2 x B92720 ..

```

```

Align seg 1/1 to: B92720 from: 1 to: 354

```

```

198 11ethg1ua1aleuglnserasmetargphcyslyslglnlth 214
|||||.....
202 ATTACAGAACACCTTCAGACACATGAGTTGTGTAAGAAACATCAAC 251
|||||.....
214 rPro1leg1uleuserleuproglythrvalmetagl 227
|||||.....
252 ACCCATGAGCTGTCTCTCCAGACGACTGATCGACACAG 291
|||||.....

```

```

seq_name: gb_est38:AV674451

```

```

seq_documentation_block:
LOCUS      AV674451      647 bp      mRNA      EST      05-OCT-2000
DEFINITION AV674451 Nori Satoh unpublished cDNA library Clona intestinalis
CDNA clone c1b15K14 5', mRNA sequence.
ACCESSION  AV674451
VERSION     AV674451.1 GI:10112450
KEYWORDS   EST.
SOURCE      Clona intestinalis.
ORGANISM   Clona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Clonidae; Clona.

```

```

REFERENCE   1 (bases 1 to 647)
AUTHORS     Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
TITLE       Expressed genes in clona intestinalis
JOURNAL     Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-Ku, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.

```

```

FEATURES
  source
    1..647
      /organism="Clona intestinalis"
      /db_xref="taxon:7719"
      /clone="c1b15K14"
      /clone_lib="Nori Satoh unpublished cDNA library"
      /tissue_type="whole animal"
      /dev_stage="tailbud"

```

```

BASE COUNT      229 a      137 c      112 g      169 t
ORIGIN

```

```

alignment_scores:
  Quality: 136.50      Length: 227
  Ratio: 1.241         Gaps: 7
Percent Similarity: 48.458 Percent Identity: 24.229

```

```

alignment block:
US-09-389-000-2 x AV674451 ..

```

```

Align seg 1/1 to: AV674451 from: 1 to: 647

```

```

144 Glulysleuargarguarlyleystyrcyscysglnleuargth 160
|||||.....
44 GAAGAGTCGACGAGAACTATTAACATTCGTGGCAAACTACGAGC 93
|||||.....
160 rleuleiProtyrValylsclYargylasnaSpala1aSerValleug 177
|||||.....
94 TTGTTCGCCGATTCAGTAACATTAAGAAAGACATGGCATCTTGCTGG 143
|||||.....
177 lu1aThrValaspTyrrValylsTyrrlearglulysIleSer..... 191
|||.....
144 AACAAACAGTTGATACATATTTATGCACACAGATGAAAGAGAC 193
|||||.....
192 ...Pro1aValMetaglInle1erhglua1aleuglnserasmetar 207
|||||.....
194 AACCCAGCTTGTAAATAAGGCTATATGTCATACCGACAAAACAGA 243
|||||.....
207 gpheCyslyslglnlthrPro1leg1uleuserleuproglythr 224
|||||.....
244 AGAAATTTGAAATTTTCGAAATCTTGAAAGAAACAAATTCAGTAACA 293
|||||.....
224 almetaglInargluanser..... 231
|||.....
294 CAAGTCGAAAGATCTAAATCTTCGACACAGCTGTCTGGGGTCAT 343
|||||.....
232 ValmetSerThrTyrrSerProgluargglyleuglnpheleuthrnsn 248
|||||.....
344 TCCTTTCAAACTACACTCCACAA..... 367
|||||.....
248 rCysTTPasnGlyCysSerThrProasp1agluserSerleuasp1ua 265
|||||.....
367 ..... 367
265 lavalArgValProSerSerAlaSerGluasnAla1leg1yAspPro 281
|||||.....
368 .....CCATCATCA.....AATGCAATGTTGATCCG 397
|||||.....
282 TyrlyThrIleSerSerAla.....AlauserIle 293
|||||.....
398 TTTATTGAACACATCATGACGACCATCTTATTGATGATCATTAATTCG 447
|||||.....
293 uasnSerleuHisThrValArgTyrrSerlyVal..... 305
|||||.....
448 AAGATGGGAATCACCCATGTCATCAGAAAGATCAAGATTCATGT 497
|||||.....
306 .....ThrProSerTyrrAspAlaThrAlaValThrAsnGluAsnIle 319
|||||.....
498 TTGCCTCACAACCTCTTTCAGTGTGTCATCAATCAGACAGTTGATA 547
|||||.....
320 SerIleHisIleu..... 323
|||||.....
548 TCATCCCATGCAAACTACAGGAGTACATCAAGATCTATCTACTATCC 597
|||||.....
324 ProSerAlaMetProProValSerSerPhe 333
|||||.....
598 TCCAGCTGCAAAATGTGCCAGTGAACCAATTT 628
|||||.....

```

```

seq_name: gb_gss28:CNS05GA7

```

```

seq_documentation_block:
LOCUS      CNS05GA7      1015 bp      DNA      GSS      26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence SP6 end of clone
025108 of library B from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  AL336040
VERSION     AL336040.1 GI:8229798
KEYWORDS   GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
            Holacanthopterygii; Acanthopterygii; Percormorpha;

```

REFERENCE	Tetraodoniformes; Tetraodontidae; Tetraodontidae; Tetraodon.
AUTHORS	1 (bases 1 to 1015) Roest-Crollius,H., Jallion,O., Dasilva,C., Fzames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1015) Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C., Bernot,A., Fzames,C., Winkler,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
TITLE	Unpublished 3 (bases 1 to 1015)
JOURNAL	Direct Submission
REFERENCE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
AUTHORS	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .
COMMENT	location/Qualifiers
FEATURES	1..1015 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone_1id="025108" /clone_1id="B" /note="Genoscope sequence ID : COAB025BE04BI-end : SP6"
BASE COUNT	231 a 293 c 271 g 204 t 16 others
ORIGIN	
alignment_scores:	Quality: 121.00 Length: 284 Ratio: 0.858 Gaps: 14 Percent Similarity: 49.648 Percent Identity: 26.408
Alignment_block:	US-09-389-000-2 x CNS05GA7 ..
Align seg 1/1 to: CNS05GA7 from: 1 to: 1015	
66 valvallystyrttyrphrthrhcyspro..... 74	
::::: :::	
167 gtgtgtcagtcagtgtggacaccagtcgccctccatccacatgaccttgtt 216	
75SerAsnThrValIyStgThGluSnaAlaTrngLyProGluGlu 89	
217 atgcataatcccccccgaataaaanacaga...gcagcttccacagagcat 263	
89 euglyleuProlLeuGlinaTgserTyTyrSerGluHisLeuGlyTyrPhePro 105	
264 CAGAGCTCAGCGCTGCAGCCACAGAGAAGCGCTGTGACTTC A.....CCT 306	
106 ThrAspLeuPheAlaCysSerGluSerleu..... 115	
307 gcagacatgcttcctcgctccacacactctccgcctcatccacatccacgat 356	
116ArgAsnGlyAsnGlyLeuGluL 123	
357 gtgcgataaaganaccatcacagtagtagacagacagagatgtttccacag 406	
123 euASnaIasertleuSerGluPheGluLylAsnLylsYlleeerleuleuL 139	
407 AGCCTGCTGGCTGCTGAGTGtGTGCTCACTGCCCTGTCGCCCTCGCCCTC 456	
140 HisSerSerLys.....GluYstLeuAArgAgGluArgIlleUsty 153	
457 CAGCTGAGGAAACCGTTGTGAAAAAGATGCCACAGATCGCATCACAG 506	
153 rCysCySGluGlnLeuAArgThrleu.....ProTYrValYsgLYA 168	

```

507 CTGCATCGACCTCAAGTCAAGTCAATCGGAAAAAGACGATCCAGAACGAGG 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 rglvysn.....AspAlaHisValIleuLualatPrValAsp 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 AACCCACATCCAGCTGAGAAAGCCGACATCTGGAGATGACC..... 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 TyValysTrpIleArgGluTyrIleSerPro,AlaValMetAlaGlnI 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 ..GTAGCTTCTCTGAGGACGACGATCGACGCGGCCCCGTGTGGC.... 643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 IethIuAlaLeuGlnSerAspMetArgPheCysIleTyrGlnGlnThr 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 .....GCANTACAGCCAGGGCTTCGCACATCGTGGAGGAATCGGCC 665
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 ProIleGluIleuSerLeuProGluThrValMetAlaGlnArgGluAsn 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 CCTTCTCTCTGACGAGTCCGACAA.....GAGGTCC 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 rValMetSerThrTyrSerProGluArgGluGlnPheLeuThrAsnT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
721 GGTACACTCTCTGACAGGCGCCACGACG.....CAGCAGCTCTCCACAG 764
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 hCysTrpAsnGlyCysSerThrProAspAlaGluSerIleuAspGlu 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
765 CCACGCGAGGCGGACACTCAGCCCGACGACCCACCTCCGACGGACAG 814
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 AlaValArgValProSerSerAlaSerGluAsnAlaIleGlyAspPr 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 .....AGCAGACACACACAGAGGGCGCCCTGTGGAGAGCC 849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 cyrIyTrpThrHisIleSerSerAlaIleuSerIleuAsnSerLeuHisT 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
850 CTGG.....TAGAACCCTGAAACCCACGAGCA 875
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 hValArgTyrTyrSerIleValThrProSerTyrAspAlaThrAlaVal 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
876 CTGTGAGACCTGAGGGGAGCTGCACCTCCATCATATGTAGANNACGTG 925
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est49:AW830314

seq_documentation_block:
LOCUS AW830314 475 bp mRNA EST 18-JUL-2000
DEFINITION sm53603.y1 Gm-cl028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl028-4926 5' similar to TR:022768 022768 HYPOTHETICAL 33.1 KD
PROTEIN: ; mRNA sequence.
ACCESSION AW830314
VERSION AW830314.1 GI:7924288
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta:
Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 475)
Shoemaker,R., Kelm,P., Vocklin,L., Erpelting,J., Correll,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
,Y., Feerson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ralster,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST project
Public Soybean EST project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)

```



```

206 targephycyslsglninglnthrpriolegluleuserleuproglyt 223
|||||
501 .....TTTGGCGTAAACTGCGGAGGCTTAAGTTGACCAAGACGGACT 546
223 hr...Valmetlaglalnrggluanservalmetserthrtysr 237
|||||
547 CCTGTGACCTGTCTCCTGACCAAGAGTAGTTACAGACGCCCTTCA 593
seq_name: gb_est72:BE548486

seq_documentation_block:
LOCUS BE548486 884 bp mRNA EST 09-AUG-2000
DEFINITION 601076062F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3461918 5',
mRNA sequence.
ACCESSION BE548486
VERSION BE548486.1 GI:9777131
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM8458 row: 1 column: 15
High quality sequence stop: 684.
Location/Qualifiers
FEATURES
Source 1. 884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3461918"
/clone_id="NIH_MGC_12"
/tissue_type="choriocarcinoma"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: pCMV-SPORT6; Site_1: NCI;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 259 a 191 c 247 g 187 t
ORIGIN
alignm_scores:
Quality: 119.00 Length: 101
Ratio: 1.776 Gaps: 3
Percent Similarity: 66.337 Percent Identity: 28.713
alignment_block:
US-09-389-000-2 x BE548486 ..
Align seg 1/1 to: BE548486 from: 1 to: 884
140 Htsserlsrlysglulysleuarglularglylelytyrcyscysgl 156
|||||
257 CATACCGAATGGAAGAGATAGAGCGCAATCGCATTTGCTGTGA 306
156 ugluleuargthrlleuLeuProtyrVallysglylrgylsasnaSPALaa 173
|||||
307 TGAGTGAATCTCTAGCGCTTGTGCAATGCC...GAGACTGACAGG 353
273 laservalleuglualathrvalaspTYrVallystYrileargglulys 189
|||||

```

```

354 COCAACTCTGACGTGACGACACAGCATTTCTGAAATACATCCAGAAAAG 403
190 lleserProalavalmetlaglalnlethrcglualaleuInserasme 206
|||||
404 CATGAGATTCCTCTTAAAAAGCAATTCAGAGCGTA..... 439
206 targephycyslsglninglnthrpriolegluleuserleuproglyt 223
|||||
440 .....TTTGGCGTAAACTGCGGAGGCTTAAGTTGACCAAGACGGACT 485
223 hr...ValmetlaglalnrggluanservalmetserthrtysrPro 238
|||||
486 CCTGTGACCTGTCTCCTGACCAAGAGTAGTTACAGACGCCCTCGATG 535
239 glu 239
|||||
536 GAG 538
seq_name: gb_est75:BE757087

seq_documentation_block:
LOCUS BE757087 449 bp mRNA EST 15-SEP-2000
DEFINITION 211433 MARC 280V Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757087
VERSION BE757087.1 GI:101171079
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovine; Bos.
REFERENCE 1 (bases 1 to 449)
AUTHORS Smith,T.P.L., Casses,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Komper,G.A., Laegreid
,W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.960904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGACGACGAGC
Plate: 62 row: 0 column: 24
Seq primer: ATTAGGTGACACTATAG.
FEATURES
Source 1. 449
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 280V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 121 a 108 c 139 g 81 t
ORIGIN
alignm_scores:
Quality: 116.00 Length: 101
Ratio: 1.731 Gaps: 3
Percent Similarity: 66.337 Percent Identity: 27.723
alignment_block:

```



```

DEFINITION      z1878408.t1 Stragatene colon (#937204) Homo sapiens cDNA clone
ACCESSION      IMAGE:511574 5', mRNA sequence.
VERSION        AA127107
KEYWORDS       AA127107.1 GI:1686468
SOURCE         EST.
ORGANISM       Homo sapiens
                human.
REFERENCE      Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
                1 (bases 1 to 437)
TITLE          Haller, J., Dierlich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
                Chissoe, S., Dierlich, N., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moo
                'M', Hultman, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
                Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, D., Trevisan, E
                Underwood, K., Wohlmann, P., Waterson, R., Wilson, R. and Merra, M.
                Generation and analysis of 280,000 human expressed sequence tags
                Genome Res. 6 (9), 807-828 (1996)
JOURNAL        97044478
MEDLINE
COMMENT        Contact: Wilson RK
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: estewatson.wustl.edu
                This clone is available royalty-free through LNM; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Insert Length: 1357 Std Error: 0.00
                Seq primer: -28M13 rev2 from AmerSham
                High quality sequence stop: 310.
FEATURES
    source
        1..437
            /organism="Homo sapiens"
            /db_xref="GDB:3844069"
            /db_xref="taxon:9606"
            /clone="IMAGE:511574"
            /clone_id="Stragatene colon (#937204)"
            /tissue="tumor"
            /cell_line="84 carcinoma cell line"
            /lab_host="SOLR cells (kanamycin resistant)"
            /note="Organ: colon; Vector: pBluescript SK-; Site_1:
            EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
            Oligo dT. T-84 colonic epithelial cell line. Average
            insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
            sequence: 5' GAATTCGCCACGAG 3' -3' adaptor sequence: 5'
            CTCGAGTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT      123 a 101 c 128 g 83 t 2 others
ORIGIN
Alignment_scores:
    Quality: 112-50 Length: 83
    Ratio: 2.045 Gaps: 2
    Percent Similarity: 66.265 Percent Identity: 30.120
alignment_block:
    US-09-389-000-2 x AA127107 ..
Align seg 1/1 to: AA127107 from: 1 to: 437
140 HisSerSerLysGluLysLeuArgArgGluArgIleLysTyrCysCysG1 156
||||| ||||| ||||| ||||| ||||| |||||
134 CATAACCAATGCAAGAAAGATGAGACGCCGCAAAATCCCATTTGCTGCA 183
||||| ||||| ||||| ||||| ||||| |||||
136 uGlnLeuArgThrLeuLeuProTyrValLysGlyArgGlySerAsnSpAlaA 173
||||| ||||| ||||| ||||| ||||| |||||
184 TGAGTTGATCTCTTAAAGTCGCGCTTGCAGATGCC...GAGACTGACAGG 230
||||| ||||| ||||| ||||| ||||| |||||
173 LasSerValLeuGluLalThrValaLysTyrValLysTyrLArgGluLys 189
||||| ||||| ||||| ||||| ||||| |||||
231 CCACGACCTCTGCAGACGACACAGATCTCCGGAATATACATCCAGGAAGA 280
||||| ||||| ||||| ||||| ||||| |||||
190 IleSerProAlaValMetAlaGlnIleThrGluLalalaLeuGlnSerAsnMe 206

```

```

281 CATGGAGATTCTCTTAAAAAGAAATTTGAGACGCA..... 316
      : : : : : : : : : : : : : : : : : : : : :
206 tArgpheCysLysLysGlnGlnThrProIleGluLeuSerIendProGly 222
      ||||| ||| : : : : : : : : : : : : : : :
317 .....TTTTCGGGAAAACGCGCCGACNNAAAAAGCTTGACACGACCGGA 361
      : : : : : : : : : : : : : : : : : : : : :

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 22:26:51 ; Search time 69.17 Seconds
(without alignments)
4958.062 Million cell updates/sec

Title: US-09-389-000-1
2128

Perfect score: 1 gaccggggggcggttggtggtf.....ccacattatgaaaaaaa 2128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	58.8	2.8	7218	1	US-08-232-463-14
2	40	1.9	3763	1	US-07-792-865D-1
3	40	1.9	7218	1	US-08-232-463-14
4	37.2	1.7	1360	3	US-08-961-083-37
5	34.8	1.6	1425	3	US-09-009-494-5
6	34.2	1.6	1002	3	US-08-576-240-1
7	34.2	1.6	1602	3	US-08-530-950-3
8	34.2	1.6	1602	3	US-08-888-429A-3
9	34.2	1.6	2924	3	US-09-357-073-1
10	34.2	1.6	2943	2	US-08-788-892-1
11	34	1.6	500	3	US-09-141-000-2
c 12	34	1.6	1544	2	US-08-837-593-1
13	33.2	1.6	747	2	US-08-679-765-2
14	33.2	1.6	747	2	US-09-196-525-2
15	33.8	1.5	3095	5	5231168-1
c 16	32.8	1.5	4032	1	US-08-107-748-3
c 17	32.8	1.5	4032	1	US-08-245-809-4
c 18	32.8	1.5	4032	4	PCT-US92-01385-3
c 19	32.6	1.5	51952	4	US-08-947-823-1
c 20	32.6	1.5	72928	3	US-09-009-913-1
c 21	32.4	1.5	2493	3	US-08-945-056-3
c 22	32.4	1.5	3827	1	US-08-170-294-6
c 23	32.4	1.5	3827	2	US-08-664-855-6
c 24	32.4	1.5	3827	2	US-08-718-751-1
c 25	32.4	1.5	3827	2	US-09-049-289-6
c 26	32.4	1.5	7399	2	US-08-418-848A-9
27	32.4	1.5	9709	2	US-08-188-583-5
28	32.4	1.5	9709	3	US-08-388-353-1

29	32.4	1.5	9709	3	US-08-488-551B-1	Sequence 1, Appli
30	32.4	1.5	15581	3	US-08-646-538-35	Sequence 35, Appl
c 31	32.2	1.5	591	1	US-08-090-523-24	Sequence 24, Appl
c 32	32.2	1.5	591	1	US-08-334-639-3	Sequence 3, Appli
c 33	32.2	1.5	591	1	US-08-398-627-24	Sequence 24, Appl
c 34	32.2	1.5	591	1	US-08-406-858-25	Sequence 25, Appl
c 35	32.2	1.5	591	4	PCT-US94-05275-25	Sequence 25, Appl
c 36	32.2	1.5	2431	3	US-08-714-918-15	Sequence 15, Appl
c 37	32.2	1.5	8791	4	PCT-US96-01735-5	Sequence 5, Appli
c 38	32.2	1.5	10564	1	US-08-206-176-5	Sequence 5, Appli
39	31.8	1.5	824	1	US-08-158-353-1	Sequence 1, Appli
40	31.8	1.5	1486	5	5516630-3	Patent No. 5516630
41	31.6	1.5	2573	3	US-08-714-918-17	Sequence 17, Appl
42	31.6	1.5	2573	3	US-08-714-918-64	Sequence 64, Appl
43	31.4	1.5	2621	2	US-08-553-619B-8	Sequence 8, Appli
c 44	31.4	1.5	5599	2	US-08-477-451-9	Sequence 9, Appli
c 45	31.4	1.5	5599	2	US-08-477-451-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,766
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTgpt-F15
US-08-232-463-14
Query Match 2.8%, Score 58.8, DB 1, Length 7218;


```

1      RESULT      3
2      US-08-232-463-14
3      ; Sequence 14, Application US/08232463
4      ; Patent No. 5670367
5      ; GENERAL INFORMATION:
6      ; APPLICANT: DORNER, F.
7      ; APPLICANT: SCHEIFLINGER, F.
8      ; APPLICANT: FALKNER, F. G.
9      ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
10     ; NUMBER OF SEQUENCES: 52
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Foley & Lardner
13     ; STREET: 1800 Diagonal Road, Suite 500
14     ; CITY: Alexandria
15     ; STATE: VA
16     ; COUNTRY: USA
17     ; ZIP: 22313-0299
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: Patentln Release #1.0, Version #1.25
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/08/232,463
25     ; FILING DATE:
26     ; CLASSIFICATION: 435
27     ; PRIOR APPLICATION DATA:
28     ; APPLICATION NUMBER: US/07/935,313
29     ; FILING DATE:
30     ; APPLICATION NUMBER: EP 91 114 300. 6
31     ; FILING DATE: 26-AUG-1991
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: BENT, Stephen A.
34     ; REGISTRATION NUMBER: 29,768
35     ; REFERENCE/DOCKET NUMBER: 30472/114 IMM
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: (703)836-9300
38     ; TELEFAX: (703)683-4109
39     ; TELEX: 899149
40     ; INFORMATION FOR SEQ ID NO: 14:
41     ; SEQUENCE CHARACTERISTICS:
42     ; LENGTH: 7218 base pairs
43     ; TYPE: nucleic acid
44     ; STRANDEDNESS: single
45     ; TOPOLOGY: linear
46     ; IMMEDIATE SOURCE:
47     ; CLONE: PTZ9pt-Fls
48     ;
49     ; US-08-232-463-14

```

Query Match	1.9%	Score 40:	DB 1:	length 7218;
Best Local Similarity	8.7%	Pred. No. 0.1:		
Matches	34:	Conservative 184:	Mismatches 174:	Indels 0:
			Gaps	0
Qy	1535	ttccatcaagctcgcgcctcagagaatgctatgtgtatccataataaacatcatcca	1594	
Db	1120	yy	1179	
Qy	1595	gtgcagcgcgtctctgtaattctctgcatcgtcagatataatctaaagtcacccctt	1654	
Db	1180	yy	1239	
Qy	1655	cctcagatgcaactgctgtaacaatcgaacatttcaattcaacttcagcatgco	1714	
Db	1240	yy	1299	
Qy	1715	cccgcgctcaagctctctccctcgcgactgcaacttgggtggcgcagcgtcctac	1774	
Db	1300	yy	1359	
Qy	1775	acatcccaactgctcgcaacagtttgggcgtattaaagcacatglttgaanttcacac	1834	

[illegible]

```

1      RESULT      4
2      US-08-961-083-37
3      ; Sequence 37, Application US/08961083
4      ; Patent No. 6159469
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Choi et. al.
7      ; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
8      ; NUMBER OF SEQUENCES: 452
9      ; CORRESPONDENCE ADDRESS:
10     ; ADDRESSEE: Human Genome Sciences, Inc.
11     ; STREET: 9410 Key West Avenue
12     ; CITY: Rockville
13     ; STATE: Maryland
14     ; COUNTRY: USA
15     ; ZIP: 20850
16     ; COMPUTER READABLE FORM:
17     ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
18     ; COMPUTER: HP Vectra 486/33
19     ; OPERATING SYSTEM: MSDOS version 6.2
20     ; SOFTWARE: ASCII Text
21     ; CURRENT APPLICATION DATA:
22     ; APPLICATION NUMBER: US/08/961,083
23     ; FILING DATE:
24     ; CLASSIFICATION: 435
25     ; PRIOR APPLICATION DATA:
26     ; APPLICATION NUMBER:
27     ; FILING DATE:
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Brookes, A. Anders
30     ; REGISTRATION NUMBER: 36,373
31     ; REFERENCE/DOCKET NUMBER: PB340P2
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (301) 309-8504
34     ; TELEFAX: (301) 309-8512
35     ; INFORMATION FOR SEQ ID NO: 37:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 1360 base pairs
38     ; TYPE: nucleic acid
39     ; STRANDEDNESS: double
40     ; TOPOLOGY: linear
41     ;
42     ; US-08-961-083-37

```

Query Match	1.7%	Score 37.2	DB 3	Length 1360
Best Local Similarity	47.8%	Pred. No. 0.27		
Matches	108	Conservative	0	Mismatches 118
				Indels 0
				Gaps 0
QY	225	agacacaacttgaatlaaatgataactagaagaacaattgtttatctcaaggagaag	284	
Db	537	AGACGATTCGTAAAAAACCGAAGAGAAGACCTAAACGAAAACACGACAGAAGATAAAG	596	
QY	285	tggaaaaaatccatgtagaactcttcagatagaactatctctatctctgtgtctatgaacgaa	344	
Db	587	TTAAAGAAAAAACCACTGTAACACCAACAACCGCCGCTCTCAAAAGCAGAAAAACCG	656	
QY	345	tggccagtgaatctcttaaacacattacttaaacagctgagaagaagaagaagactcttg	404	
Db	657	CTCCAGCTCCAAACCAACGAGATCCAGCTGGAACCAACCAAGAAAAACCAAGCTGATC	716	
QY	405	aaagtcaggatgaattactatctactgaacttaaaccttggaacaagaatcaaa	450	

Db 717 AACAGCTGAGAGACTATGCTGTAGATGACAGAGAAATATTA 762

RESULT 5

US-09-009-494-5
Sequence 5, Application US/09009494
Patent No. 6150340

GENERAL INFORMATION:

APPLICANT: Black, Michael T.
APPLICANT: Lawlor, Elizabeth J.

TITLE OF INVENTION: No. 6150340e1 Compounds
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia
STATE: PA

COUNTRY: USA
ZIP: 19103-2793

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/009,494
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/037,857
FILING DATE: 07-FEB-1997

APPLICATION NUMBER: 60/044,365
FILING DATE: 28-APR-1997

APPLICATION NUMBER: 60/044,366
FILING DATE: 28-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GM50028
TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1425 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

US-09-009-494-5

Query Match 1.6%; Score 34.8; DB 3; Length 1425;
Best Local Similarity 47.3%; Pred. No. 1.4;

Matches 105; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Db 771 ggagctgaagccatcaatgaatttagatttggcaaaaagaataacattcctt 830

Db 3 GCATTTTGAAACAGTTAAGAGCTTGAAGTTCACGTAGATTAAAGCGACTCAAAAT 62

Db 831 ttttttaataccctgaataatttaaggttgatattcaggcgatggaatgatctgc 890

Db 63 GTTTTTCCTCACCAGCGCATTTTGGAGCAGAACCTTAACCTCAAAATGATTATCGA 122

Db 891 tttaactgaaccactgaatgaataaataatgaatgtgttaataatactggacacatg 950

Db 123 CTGAGCAATATCCAGGTGCTTACACAGTGTAAATGAAGGTGAGAGATGCGCAATGCG 182

Db 951 tcctcaaacactgtttaagactgaataacgcaactggtgacctga 992

Db 183. TGCTGCAATGGCAGCTAAATATGAAATGCAACAGAAATCTTA 224

RESULT 6

US-08-576-240-1
Sequence 1, Application US/08576240

Patent No. 6074862

GENERAL INFORMATION:

APPLICANT: Stein, Bernd
APPLICANT: Yang, Maria

TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/576,240
FILING DATE: 20-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.403
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1002 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

FEATURE:
NAME/KEY: CDS
LOCATION: 1..1002

US-08-576-240-1

Query Match 1.6%; Score 34.2; DB 3; Length 1002;
Best Local Similarity 49.7%; Pred. No. 1.8;

Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Db 371 cttaacagctgaagaagaagaagctcttgaagctcaagtgaaataactatgcctt 430

Db 799 CTCAAAACAGGTGTAGAGAGACCATCCGACACACTCCAGCAGACAAAGTTCTGCGAG 858

Db 431 aaactggaacaagaataaagcttaccagaagaatcaacaatgaacgacctatactta 490

Db 859 TTTGTTGACTTTACTCTCAGATGCTTAAAGAAAGATTCAAAGAACGGCTTACATACCA 918

Db 491 gctgaatgttccaggggtctggtttacatcaagttcttaaaagcaacaggtg 545

Db 919 GAGCTAATGCACAATCATCTTTTCACCTTACATGATCAATCAAGAAAGAGATGTGG 973

RESULT 7

US-08-550-950-3
Sequence 3, Application US/08530950

Patent No. 5736381

GENERAL INFORMATION:

APPLICANT: Davis, Roger J.
APPLICANT: Raingeaud, Joel

APPLICANT: Gupta, Shashi
APPLICANT: Derijard, Benoit

TITLE OF INVENTION: CYTOKINE-, STRESS-, AND

TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
TITLE OF INVENTION: KINASES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,950
FILING DATE: 19-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1602 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-530-950-3

Query Match 1.6%; Score 34.2; DB 1; Length 1602;
Best Local Similarity 49.7%; Pred. No. 2.3;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 371 cttaaacagctagaagaagaagaagactcttgaaagtcagtgaaatctatgcactt 430
DB 1042 CTCGAACAGCTGTAGAGAGCCATCCGACACTCCAGCAGCAAGTCTCTGCGAG 1101
QY 431 aaactggaacaagaatcaaggtcttaccagaagaatcaaatgagcgctacataccta 490
DB 1102 TTGTGTGACTTTACCTCAGAGTGGCTTAAGAGAGATTCGAAGACGCGCTACATACCA 1161
QY 491 gctgaatgtctcaggggtcttggtttacatcaagttcttaaaagcaagcggtgg 545
DB 1162 GAGCTAATGCAACATCATTTTTCACCCCTACATGAATCCAAAGAGACAGATGTGG 1216

RESULT 8
US-08-888-429A-3
Sequence 3, Application US/08888429A
Patent No. 6136596

GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Whitmarsh, Alan
APPLICANT: Tournier, Cathy
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/888,429A
FILING DATE: 07-JUL-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/530,950
FILING DATE: 19-SEP-1995
APPLICATION NUMBER: 08/446,083
FILING DATE: 19-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/053001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 299354
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 244...1245
US-08-888-429A-3

Query Match 1.6%; Score 34.2; DB 3; Length 1602;
Best Local Similarity 49.7%; Pred. No. 2.3;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 371 cttaaacagctagaagaagaagaagactcttgaaagtcagtgaaatctatgcactt 430
DB 1042 CTCGAACAGCTGTAGAGAGCCATCCGACACTCCAGCAGCAAGTCTCTGCGAG 1101
QY 431 aaactggaacaagaatcaaggtcttaccagaagaatcaaatgagcgctacataccta 490
DB 1102 TTGTGTGACTTTACCTCAGAGTGGCTTAAGAGAGATTCGAAGACGCGCTACATACCA 1161
QY 491 gctgaatgtctcaggggtcttggtttacatcaagttcttaaaagcaagcggtgg 545
DB 1162 GAGCTAATGCAACATCATTTTTCACCCCTACATGAATCCAAAGAGACAGATGTGG 1216

RESULT 9
US-09-357-073-1
Sequence 1, Application US/09357073
Patent No. 6033910

GENERAL INFORMATION:
APPLICANT: Brett P. Monia
APPLICANT: Lex M. Cowsett
TITLE OF INVENTION: ANTISENSE MODULATION OF MAP KINASE KINASE 6 EXPRESSION
FILE REFERENCE: RTS-0086
CURRENT APPLICATION NUMBER: US/09/357,073
CURRENT FILING DATE: 1999-07-19
NUMBER OF SEQ ID NOS: 47
SEQ ID NO 1
LENGTH: 2924
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (341)...(1345)
US-09-357-073-1

Query Match 1.6%; Score 34.2; DB 3; Length 2924;
Best Local Similarity 49.7%; Pred. No. 3.2;


```

1 APPLICANT: Au-Young, Janice
2 APPLICANT: Goli, Surya K.
3 APPLICANT: Hillman, Jennifer L.
4 TITLE OF INVENTION: A NOVEL HUMAN UBIQUITIN-CONJUGATING ENZYME
5 NUMBER OF SEQUENCES: 5
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Incyte Pharmaceuticals, Inc.
8 STREET: 3174 Porter Drive
9 CITY: Palo Alto
10 STATE: CA
11 COUNTRY: U.S.
12 ZIP: 94304
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Diskette
16 COMPUTER: IBM Compatible
17 OPERATING SYSTEM: DOS
18 SOFTWARE: FASTSEQ Version 1.5
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/679,765
22 FILING DATE: Filed Herewith
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Billings, Lucy J.
25 REGISTRATION NUMBER: 36,749
26 REFERENCE/DOCKET NUMBER: PF-0093 US
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: 415-855-0555
29 TELEFAX: 415-845-4166
30
31 INFORMATION FOR SEQ ID NO: 2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 747 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: cDNA
38 IMMEDIATE SOURCE:
39 LIBRARY: PGANOT01
40 CLONE: 61887
41
42 US-08-679-765-2
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
9
```

```
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,525
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/679,765
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0093 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: PGANNO701
CLONE: 61887
US-09-196-525-2
```

```
Query Match 1.6%; Score 33.2; DB 2; Length 747;
Best Local Similarity 53.0%; Pred. No. 2.9;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
QY 195 caagaagtagtgaaatgaaacccgtatgagacacacttgaaatgaatgatgagactag 254
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 CAAGCAGTAGTGAGGAGAGTCCGATGAGATCAAGTGAAGTGTTCAGCAAGACAG 224

QY 255 aaaaagcaatgttatctcaaggagaaagtcgaaaaaaatccatggaactcttcagata 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 AAAGAGAACAGTTCAGCCCAAGAAAGAGGAGAAATATCCAGCAAAACCGTGCTA 284

QY 315 gactatcttctatt 328
| | | | |
DB 285 AATTGTCAACTAGT 298
```

```
RESULT 15
5231168-1
; Patent No. 5231168
; APPLICANT: DIEGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN;
; VUUST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:1:
; LENGTH: 3095
5231168-1
```

```
Query Match 1.5%; Score 32.8; DB 5; Length 3095;
Best Local Similarity 47.1%; Pred. No. 8.8;
Matches 132; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

QY 165 cctgaacagcagttgctggaagaagtcgcgaagaagtcagtgaaatggaanaa-cccgltat 223
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1318 -catgaatagtagaagtgaagaattctacagaagaataaaatgaaaaaggtcaacat 1377
```

```
QY 224 gagacacacttgaaatgaatgagactagaaaagcaattgttatccaaagagaaa 283
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1378 gaaatagtagaggttgagaagaattctaccagaagataaaaatgaaaagtgaacagaa 1437

QY 284 gtgaaaaaalcatalggaactcttcagatagactatctctatctgltctatgaaaga 343
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1438 atagtagaagttgagaagaattctaccagaagataaaaatgaaaaggtcaacatgaaata 1497

QY 344 atgccagttgaaatccttaaacacattacttaaacagctgagaagaagaagaactctt 403
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1498 gttagagttgagaagaattctaccagaagataaaaatgaaaagttcaacatgaaatagta 1557

QY 404 gaaagtcagtgaaatactatgcacttaaacctggaacag 443
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1558 gaagttgaagaattctaccagaagataaaaatgaaaaag 1597
```

Search completed: January 19, 2001, 01:45:12
Job time: 11901 sec

THIS PAGE BLANK (USPTO)


```

XX WPI: 2000-237872/20.
DR P-PSDB: Y79269.
XX
PT Testis specific Helix Loop Helix proteins expressed in cancers and
PT useful for the prevention, diagnosis and treatment of prostate, bladder
PT and ovarian tumors -
XX
PS Claim 5; Fig 2A-D; 62pp; English.
XX
CC This sequence is that of cDNA encoding human PHELIx (see Y79269),
CC a novel basic Helix Loop Helix protein thought to act as a
CC transcription factor. PHELIx normally exhibits a testis-specific
CC expression pattern but is up-regulated in prostate and other types
CC of cancer. The cDNA clone, termed GPRC12 (ATCC 98956), was
CC isolated from a normal testis cDNA library using a suppression
CC subtractive hybridization method. The gene maps to chromosome
CC 13p13.1-13.3. The invention provides diagnostic and therapeutic
CC methods useful in the management of various cancers which express
CC PHELIx, including prostate cancer, bladder cancer, ovarian cancer
CC and testicular cancer, including therapies aimed at inhibition the
CC transcription, translation, processing or function of PHELIx. The
CC expression pattern of PHELIx suggests that is an ideal target for a
CC cancer vaccine approach to prostate cancer.
XX
SQ Sequence 2128 BP; 675 A; 463 C; 464 G; 526 T; 0 other:

Query Match      100.0%; Score 2128; DB 21; Length 2128;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaccgggggggggttggttcacccgctcgtgcgtactggtctcttggtggcccttaa 60
DB 1 gaccgggggggggttggttcacccgctcgtgcgtactggtctcttggtggcccttaa 60
QY 61 tgtctgtgctctaagtgctgaggggaaagcgggagggtctctggtcctgacactag 120
DB 61 tgtctgtgctctaagtgctgaggggaaagcgggagggtctctggtcctgacactag 120
QY 121 aaggaagagagaactaacactcgagcgtgtgagaccacacgcctgaaacagcagttg 180
DB 121 aaggaagagagaactaacactcgagcgtgtgagaccacacgcctgaaacagcagttg 180
QY 181 ctggagaagagccggaagaagtagtgaatggaacccgttatgtgagacacactggaat 240
DB 181 ctggagaagagccggaagaagtagtgaatggaacccgttatgtgagacacactggaat 240
QY 241 aatgatgaactagaaaagcaaatgtttatctcaaggagaagaagtgaaaaaatccatg 300
DB 241 aatgatgaactagaaaagcaaatgtttatctcaaggagaagaagtgaaaaaatccatg 300
QY 301 aaactcttcagatagactatcttctatctcgtgtctatgaagaaatgccaagtggaa 360
DB 301 aaactcttcagatagactatcttctatctcgtgtctatgaagaaatgccaagtggaa 360
QY 361 aaacacattacttaaacagctcagaagaagaagaagactctgaaagtccaagtgaata 420
DB 361 aaacacattacttaaacagctcagaagaagaagaagactctgaaagtccaagtgaata 420
QY 421 ctatgcaacttaactgtaacaagaatcaaaaggtctacacagaagatcaacaatgaa 480
DB 421 ctatgcaacttaactgtaacaagaatcaaaaggtctacacagaagatcaacaatgaa 480
QY 481 taataactagctgaatgtctcaggttctggtttacatcaagttcttaaaaggcaaca 540
DB 481 taataactagctgaatgtctcaggttctggtttacatcaagttcttaaaaggcaaca 540
QY 541 ggtggaatcaactgctcagaatgcaagaagaatctagtgaaaacgcaaaaatgacatcta 600
DB 541 ggtggaatcaactgctcagaatgcaagaagaatctagtgaaaacgcaaaaatgacatcta 600
QY 601 ttaagtgaagatgtcacgtgtggtcactggtcgtactgtactgtacagaataatttcaaac 660

```

```

DB 601 ttaagtgaagatgtcacgtgtggtcactggtcgtactgtactgtacagaataatttcaaac 660
QY 661 atagcagaagatcacccatcacatcagtgacagaagaagcagcagcctttgtagat 720
DB 661 atagcagaagatcacccatcacatcagtgacagaagaagcagcagcctttgtagat 720
QY 721 tgcataatcaacatggttctcttgaagtgctccttcactaaatgcccgaaggagctgaa 780
DB 721 tgcataatcaacatggttctcttgaagtgctccttcactaaatgcccgaaggagctgaa 780
QY 781 gccataagtttaattagatttgcgaaaaagaataacacattacacgtttgtttata 840
DB 781 gccataagtttaattagatttgcgaaaaagaataacacattacacgtttgtttata 840
QY 841 atccctgaaatttaaaagttgatttcaagggacatgaaatgataatgcttaactgaa 900
DB 841 atccctgaaatttaaaagttgatttcaagggacatgaaatgataatgcttaactgaa 900
QY 901 ccactgcaaatggaaaaaatgagtaatgtgttaaatatctgacacatgctccctcaaac 960
DB 901 ccactgcaaatggaaaaaatgagtaatgtgttaaatatctgacacatgctccctcaaac 960
QY 961 actgttaagactgtaaaaacgcacactgggctggaagaacttgatggccctgcaaggctcc 1020
DB 961 actgttaagactgtaaaaacgcacactgggctggaagaacttgatggccctgcaaggctcc 1020
QY 1021 taacagcgaacacccgtgatatatttccctactgatactatttgcctctgtaactttaag 1080
DB 1021 taacagcgaacacccgtgatatatttccctactgatactatttgcctctgtaactttaag 1080
QY 1081 aatgcaaatgggttgtaattaatgcttctgttcagaagttcgaagaatacaaaaagatc 1140
DB 1081 aatgcaaatgggttgtaattaatgcttctgttcagaagttcgaagaatacaaaaagatc 1140
QY 1141 tctcttctcatcaagaagaaataaactaaagaaggaaagaaatcaaatatgctgtgag 1200
DB 1141 tctcttctcatcaagaagaaataaactaaagaaggaaagaaatcaaatatgctgtgag 1200
QY 1201 cagctggtactctcttcgtccgtatgttaaaagggaagaagatgtagtgcctcagttctt 1260
DB 1201 cagctggtactctcttcgtccgtatgttaaaagggaagaagatgtagtgcctcagttctt 1260
QY 1261 gagcacaagttgtatgtgttaaatatataccggagagaagaatctccagccgttaaggcc 1320
DB 1261 gagcacaagttgtatgtgttaaatatataccggagagaagaatctccagccgttaaggcc 1320
QY 1321 cagaattacagaagcactcagaagcaatgaggtttgtgaagaacaacaacaccatc 1380
DB 1321 cagaattacagaagcactcagaagcaatgaggtttgtgaagaacaacaacaccatc 1380
QY 1381 gagctgtctctcccaaggcacctgtaigtgcaagcgggaaacagtgtagagacattac 1440
DB 1381 gagctgtctctcccaaggcacctgtaigtgcaagcgggaaacagtgtagagacattac 1440
QY 1441 tcccttgaagagagggctccaattcctactaataacgctgctggaatgggtgtccactcct 1500
DB 1441 tcccttgaagagagggctccaattcctactaataacgctgctggaatgggtgtccactcct 1500
QY 1501 gatcagaagagctcccttgatgaaagctgtgagagttccatcaagctccgcctcagaagat 1560
DB 1501 gatcagaagagctcccttgatgaaagctgtgagagttccatcaagctccgcctcagaagat 1560
QY 1561 gctatggtgatacatataaactcaacttccagtgacagcgtctgtctgtaattctgtt 1620
DB 1561 gctatggtgatacatataaactcaacttccagtgacagcgtctgtctgtaattctgtt 1620
QY 1621 catactgtcaagataatttctaaagtcacccctccctacagatgaaactgtgttaacaat 1680
DB 1621 catactgtcaagataatttctaaagtcacccctccctacagatgaaactgtgttaacaat 1680
QY 1681 cagaacatttcaatttacttcaacttcaagcattgcccgggtctcaagcttctccctggc 1740

```


D _b	1661	cagaaacatttcaattcaattcaattcaacttcaagcagatgccccgggtctcaagcttctccctggc	1740
Q _y	1741	actgcaactctgtgtgttgggccagagctgtgacacacacaccccaactgctgtgcaacagtttt	1800
D _b	1741	actgcaactctgtgtgttgggccagagctgtgacacacacaccccaactgctgtgcaacagtttt	1800
Q _y	1801	gggggtattaaagcacatgttttgaattacacgtctcaaccacactgagggcgagttt	1866
D _b	1801	gggggtattaaagcacatgttttgaattacacgtctcaaccacactgagggcgagttt	1866
Q _y	1861	gacacatctaggaaaagtgcgacacaagaagatgttttgaagctctcaccccaagaccataa	1920
D _b	1861	gacacatctaggaaaagtgcgacacaagaagatgttttgaagctctcaccccaagaccataa	1920
Q _y	1921	tcaaaagagatgtgcgatgttttgctgctctgtataataatgcaactcaaaagctcttcgcgataga	1980
D _b	1921	tcaaaagagatgtgcgatgttttgctgctctgtataataatgcaactcaaaagctcttcgcgataga	1980
Q _y	1981	agaccacagcgaaaagaagctgycgcacacacatgacatctcaattcaacacacttgaagcc	2040
D _b	1981	agaccacagcgaaaagaagctgycgcacacacatgacatctcaattcaacacacttgaagcc	2040
Q _y	2041	ccggccagccagagagctcacaagaacaatgacctcagtgacatacctctctttctcaa	2100
D _b	2041	ccggccagccagagagctcacaagaacaatgacctcagtgacatacctctctttctcaa	2100
Q _y	2101	aaaatattcccaattatgataaaaaa	2128
D _b	2101	aaaatattcccaattatgataaaaaa	2128

RESULT	2	
ID	T05868	
ID	T05868	standard; DNA; 3399 BP.
AC	T05868;	
XX		
DT	14-AUG-1996	(first entry)
XX		
DE	Chicken leucocytozoan DNA encoding immunogenic protein for vaccines	
XX		
KW	Chicken leucocytozoan; immunogen; recombinant vaccine; protection;	
KW	immunisation; vaccination; ss.	
XX		
OS	Chicken leucocytozoan.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..3399
FT		/*tag= a
FT	misc_feature	1150..3218
FT		/*tag= b
FT		/note= "fragment referred to in the Claims, for
FT		use as insert in a recombinant vaccine
FT		against chicken leucocytozoan disease"
XX		
XX	JP07284392-A.	
XX		
XX	31-OCT-1995.	
XX		
XX	19-APR-1994;	94JP-0080643.
XX		
XX	19-APR-1994;	94JP-0080643.
XX		
PA	(DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.	
PA	(KITA) KITASATO KENKUSHO SH.	
XX		
XX	WPI: 1996-006311/01.	
DR	P-PSDB; R97866.	
XX		
PT	Chicken leucocytozoan immunogenic protein - used in a recombinant	
PT	vaccine against chicken leucocytozoan disease	
XX		
PS	Claim 6; Page 6-9; 35pp; Japanese.	

CC T05866 encodes a chicken leucocytozoan immunogenic protein, this DNA
 CC or a fragment of it can be used in a recombinant vaccine to immunise
 CC against chicken leucocytozoan disease. The DNA is used in a vector
 CC and operatively linked to an expression regulatory sequence as in
 CC standard practice.
 XX
 S0 Sequence 3399 BP, 1577 A, 508 C, 798 G, 516 T, 0 other;

Query Match	2.3%	Score 48;	DB 17;	Length 3399;
Best Local Similarity	48.2%;	Pred. No. 0.0014;		
Matches 164;	Conservative 0;	Mismatches 175;	Indels 1;	Gaps 1.

Oy	168	gaacagcagttgctggaagaagtcgcgaagaagtgatggaagaaacccg- ttatgag	226
Db	2227	gaagaaagagaaagaaagaaacaatgaagaagaagaagaaagtaacatgaagaagaa	2286
Oy	227	aacgaacttgatlaaatgtgactgacagaagaacaaatggttacttcgaagagaagt	286
Db	2287	aaagaagaagatatactactgagaagaagaaagaagaagtaaccatggaagaagaaagaa	2346
Oy	287	gaagaaatccatggaaccttcacgaatagactacttctatctggtctatgacgaatg	345
Db	2347	gaagtaacacatgagaagaagaagaagaagtaacacatgaagaagaagaagaagta	2406
Oy	347	ccagtggaatccttaaacacatacttaaacgctgagaagaagaagaagctcttga	406
Db	2407	acacatgagaagaagaagaagaagtaacacatggaacgaagaagaagaagtaacacatga	2466
Oy	407	agtaagatgaatatactatgacttaacccggaacagaatcctaaggttaccagaagtc	466
Db	2467	gaagaaagaaagaaagatgaacatactgaagaagaagaagaagaagtaacacatgaagaagaa	2526
Oy	467	aacatgaagccgtacataccttagtggaagtgtccag	506
Db	2527	gaaagatcaacatgagaagaagaagaagaagtaacacatg	2566

RESULT	3	
ID	X20250/c	
	X20250 standard; DNA; 111309 BP.	
XX		
AC	X20250;	
XX		
DT	04-MAY-1999 (first entry)	
XX		
DE	Borrelia burgdorferi polynucleotide sequence #3.	
XX		
KW	Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;	
KW	epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;	
KW	infection; diagnosis; characterisation; detection; ds.	
OS	Borrelia burgdorferi.	
XX		
FN	W09858943-A1.	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998; 98WO-US12764.	
XX		
PR	03-SEP-1997; 97US-0057483.	
PR	20-JUN-1997; 97US-0050359.	
PR	22-JUL-1997; 97US-0053344.	
PR	22-JUL-1997; 97US-0053377.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(MEDI-) MEDIMQNE INC.	
PI	Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;	
PI	White OR;	
DR	WPI; 1999-081217/07.	

Query Match	1.8%	Score 38.8	DB 20	Length 3029	Mismatches 112	Conservative 0	Mismatches 122	Indels 0	Gaps 0
Best Local Similarity	47.9%	Pred. No. 0.59							
Matches 112: Conservative	0	Mismatches 122	Indels 0	Gaps 0					
237 aattaatgatagaatagaaagcaatggttactctcaagagaagaatgysaaatcc	296								
3019 ATTGAATTTCTTAACCACTTACCAAAAAATTTGATTAACCAAAAGAAATTAAGATTAC	2960								
297 atggaactcttcagatagactactctctatctcgctgctatgaacgaatgccagtggat	356								
2959 AATAAGATTTTGGTACATATTATTTGATTTGAAATTAATAAGAAAAACAATAAT	2900								
357 ccttaaacacattacttaataacgcctagaagaagaagaagctcttgaagctcaatga	416								
2899 TCTTTAACACAGAAATTTTAACTAAAGGCTATATGAGGATTTAAATATGGAAGTGC	2840								
417 aatactacacttaactaaactgacagaagaatacaagcttaccagaagatcaaca	470								

Db 2839 AATGACGCTATTGAACTATATAAAAAGGCCAATATGTAAGTTGATGTCACAA 2786

RESULT 10

XX V74463

XX V74463 standard; DNA: 7953 BP.

XX V74463;

XX 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #152.

XX

XX Computer readable medium; vaccine; *S. aureus* infection; immunodetection;

XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX skin infection; surgical wound infection; scalded skin syndrome;

XX toxic shock syndrome; ds.

XX

OS Staphylococcus aureus.

XX

XX

XX Key Location/Qualifiers

XX misc-feature 1081..1140

XX /tag=a

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX

XX misc-feature 2881..2940

XX /tag=b

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX

XX misc-feature 4681..4740

XX /tag=c

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX

XX misc-feature 6481..6540

XX /tag=d

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX

XX EP786519-A2.

XX 30-JUL-1997.

XX

XX 07-JAN-1997; 97EP-0100117.

XX

XX 05-JAN-1996; 96US-0009861.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

XX Rosen CA;

XX

XX WPI; 1997-374922/35.

XX

XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*

XX stored on computer readable medium and used in the production of

XX anti-*S. aureus* vaccines

XX

PS Claim 1; Page 770-774; 3271pp; English.

XX

XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences

XX of the invention. The DNA sequences are recorded on a computer readable

XX medium, preferably selected from a floppy or hard disk, random access

XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

XX the *S. aureus* DNA sequences allows putative functions to be assigned so

XX that protein-encoding or regulatory regions of commercial, therapeutic or

FT	CDS	complement (46268..47422)
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-19831 and
FT	FT	MORF-20106, the encoded protein shows 43.20
FT	FT	percentage identity to glycerol-3-phosphate
FT	FT	dehydrogenase (GUT2) from <i>S. cerevisiae</i> "
FT	FT	49377..49643
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"The encoded protein shows 48.86 percentage
FT	FT	identity to phosphotransferase (ptsH) from <i>Mycobacteria</i>
FT	FT	<i>capricolum</i> "
FT	FT	50060..51520
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-19832 and
FT	FT	MORF-20108, the encoded protein shows 41.92
FT	FT	percentage identity to spermidine/
FT	FT	putrescine transport ATP-binding protein
FT	FT	(potA) from <i>E. coli</i> "
FT	FT	51525..52382
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20110, the
FT	FT	encoded protein shows 26.51 percentage
FT	FT	identity to spermidine/putrescine transport
FT	FT	system permease protein (potB) from <i>E. coli</i> "
FT	FT	52366..53220
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20111, the
FT	FT	encoded protein shows 29.45 percentage
FT	FT	identity to spermidine/putrescine transport
FT	FT	system permease protein C (potC) from <i>E. coli</i> "
FT	FT	54658..55605
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20112, the
FT	FT	encoded protein shows 36.60 percentage
FT	FT	identity to stialoglycoprotease (gcp)
FT	FT	from <i>Pasteurella haemolytica</i> "
FT	FT	complement (56970..58310)
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-19834,
FT	FT	MORF-20114 and MORF-20115, the encoded protein
FT	FT	shows 43.02 percentage identity to signal
FT	FT	recognition particle protein (ffh) from <i>B.</i>
FT	FT	<i>subtilis</i> "
FT	FT	58117..59079
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20114 and
FT	FT	MORF-20115, the encoded protein shows 44.78
FT	FT	percentage identity to purine-nucleoside
FT	FT	phosphorylase (deod) from <i>E. coli</i> "
FT	FT	59083..59754
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20117, the
FT	FT	encoded protein shows 83.03 percentage
FT	FT	identity to deoxyribose-phosphate aldolase
FT	FT	(deoc) from <i>Mycobacteria pneumoniae</i> "
FT	FT	complement (64898..65731)
FT	FT	/*tag=
FT	FT	/label=
FT	FT	/note=
FT	FT	"Previously identified as MORF-20122, the
FT	FT	encoded protein shows 30.25 percent
FT	FT	identity to the protein disclosed in
FT	FT	GB:D26185..99 from <i>B. subtilis</i> "
FT	FT	complement (65713..66249)

Query Match	Best Local Similarity	Score	DB	Length
Matches 139; Conservative	45.4%	167;	Indels	0; Gaps
QY 262	aattggtatctcagaagagaagtgagaaatactcgaactctcagatagactac	321		
DB 261143	agttgaaacaacaagaagatattgagcaaaaaactccaagatttcccaacttca	261202		
QY 322	ttctattcggtctcatgaacgaatagccagtggaatccttaaacactttcta	381		
DB 261203	gaacgcttctcctaatttagctgtgaacgaggagatgtgaatlaagtgaa	261262		
QY 382	agaagaagaagaagagactcttgaagtgcaagtgaatactatgcactta	441		
DB 261263	agaagcaacttgaagaagcaactggttaaatgtgttaataatgcttgaag	261322		
QY 442	agaatcaaaaggtcttaccagaagatcaacaatgaaacgcytacatc	501		
DB 261323	cttagacaactcaacacatcacttaaccagaagcagatgtcaaacagcca	261382		
QY 502	tcaaggtctcggtttacatcaagtcttcaaaagccaacagtgatcact	561		
DB 261383	aaggttctgctgatttcacatgatgttcataaaaaactcaatgagatga	261442		
QY 562	gcaaga 567			
DB 261443	acaaa 261448			
RESULT 12	V52331/c			
ID	V52331 standard; DNA; 6846 BP.			
AC	V52331;			
DE	23-OCT-1998 (first entry)			
XX	Streptococcus pneumoniae genome fragment SEQ ID NO:198.			
XX	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;			
XX	computer readable medium; vaccine; pharmaceutical composition; ds.			
XX	Streptococcus pneumoniae.			
XX	MO981B931-A2.			
XX	07-MAY-1998.			
XX	30-OCT-1997; 97MO-US19588.			
XX	31-OCT-1996; 96US-0029960.			

DE	DNA encoding CbpA of serotype 4.
XX	
RW	Choline binding protein A; CbpA; truncate; Immune response; infection;
KW	pneumococcal bacterium; vaccine; ss.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	W09951187-A2.
PD	14-OCT-1999.
XX	
PE	07-APR-1999; 99WO-US07668.
XX	
PR	07-APR-1998; 98US-0056019.
PR	07-APR-1998; 98US-0080878.
XX	
PA	(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX	
PI	Tuomanen EI, Masure HR;
XX	
DR	WPI: 1999-620161/53.
DR	P-PSDB; Y49226.
XX	
PT	Novel polypeptides, used to develop products for the diagnosis,
PT	prevention and treatment of pneumococcal infections
XX	
PS	Disclosure: Page 9-10; 85pp; English.
XX	
CC	The invention provides novel isolated polypeptides comprising the amino
CC	acid sequence of an N-terminal choline binding protein A (CbpA) truncate
CC	(CBPA). The polypeptides can be selected from sequences shown in Y49225,
CC	Y49227, Y49230, Y49231, Y49233 and Y49238. The polypeptides can be used
CC	for inducing an immune response in a subject which has been exposed to or
CC	infected with a pneumococcal bacterium. They can also be used for
CC	preventing infection by a pneumococcal bacterium. Vaccines comprising
CC	the polypeptides or encoding nucleic acids can be used for treating a
CC	subject infected with or exposed to a pneumococcal bacterium. Antibodies
CC	specifically binding the polypeptides can be used for detection and
CC	diagnosis and for preventing pneumococcal attachment to a mucosal
CC	surface. The products can be used in humans and other animals such as
CC	domestic animals, such as feline or canine subjects, farm animals such as
CC	bovine, equine, caprine, ovine, and porcine subjects, wild animals
CC	(whether in the wild or in a zoological garden), research animals, such
CC	as mice, rats, rabbits, goats, sheep, pigs, dogs, cats, i.e. for
CC	veterinary medical use.
XX	
SO	Sequence 1969 BP; 792 A; 371 C; 459 G; 347 T; 0 other;
	Query Match 1.7%; Score 37.2; DB 20; Length 1969;
	Best Local Similarity 47.8%; Pred. No. 1.3;
	Matches 108; Conservative 0; Mismatches 118; Indels 0; Gaps 0
OY	225 agacacaacttgaatataatgatgaactggaagaagcaaatgtttatctaagagagaag 284
DB	1143 agacagatgtataaaaaagcaagaagaagctaaagcaagaagaagcagaagataaag 1202
OY	285 tggaaaaaaatccatggaactcttcagatagactatcttattctgtgtatagaaagaa 344
DB	1203 ttaaagaanaaacagcttgaaacaacacaaacagcgccggttccaaaagcagaaaaacag 1262
OY	345 tggcagctggatctcttaaacacattacttaaacagctggaagaagaagaagaagctctg 404
DB	1263 ctccagcttccaaaacacagagatcccgctggaacacaaagcaagaagaacacagctgac 1322
OY	405 aaagcgaagtgaatatctatgacctaataactggacaagaatcaaa 450
DB	1323 aacaagctgaaagaagactatgtctgtagatcagaagaagaatataa 1368

XX	Z34452.
AC	01-FEB-2000 (first entry)
XX	DNA encoding choline binding protein A (Cbpa).
DT	Choline binding protein; cbpa; adhesin; immunogen;
XX	vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KM	otitis media; pneumonia; ss.
KW	Streptococcus pneumoniae.
OS	
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	1..1969
FT	/tag= a
FT	/transl_except= (pos:1030..1033, aa:Ieu)
FT	/note= "apparent 1 base insertion, alters the reading frame"
XX	
PX	WO9951188-A2.
XX	
PD	14-OCT-1999.
XX	
PF	07-APR-1999; 99WO-US07669.
XX	
PR	07-APR-1998; 98US-0056019.
XX	07-APR-1998; 98US-0080878.
XX	
PA	(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA	(MED-) MEDIMUNE INC.
XX	
PI	Tuomanen EI, Masure HR, Witzemann TM, Johnson LS, Koenig S;
XX	WPt: 1999-633690/54.
DR	P-PsDB: Y32099.
XX	
PT	New N-terminal choline binding protein A truncate polypeptides, used to
PT	develop products for the diagnosis, prevention and treatment of
PT	pneumococcal infections -
PS	Disclosure: Page 110-111; 160pp; English.
XX	
CC	This DNA sequence codes for the choline binding protein A (cbpa)
CC	of Streptococcus pneumoniae serotype type 4. The invention provides
CC	a novel N-terminal CbpA truncated polypeptides (see Y32098-110 and
CC	Y32179-90) and polynucleotides encoding them, host-vector systems,
CC	and antibodies that specifically bind to the polypeptides. The
CC	invention also relates to vaccines including the polypeptides, to
CC	which provide protection or elicit protective antibodies to
CC	bacterial infection, specifically pneumococcus, and to antibodies
CC	and antagonists against such polypeptides for use in diagnosis and
CC	passive immunotherapy. The polypeptides and/or polynucleotides are
CC	also useful as competitive inhibitors of bacterial adhesin of
CC	pneumococcus.
XX	
SQ	Sequence 1969 BP; 792 A; 371 C; 459 G; 347 T; 0 other:
Query Match	1.7%; Score 37.2; DB 20; Length 1969;
Best Local Similarity	47.8%; Pred No.1.3; Indels 0; Gaps 0;
Matches 108; Conservative	0; Mismatches 118;
OY	225 agaccacacttgaaatgaatagtgactagaagcaaatgtttattctcaggsgaaag 284
DB	1143 agacgatcgtaaaaaaacagagaagaagcctaacaagaacagacagagaagttaaag 1202
OY	288 tggaaaataatccatggaactcttcagatagctatcttcttcgtgcctctgaagaa 344
DB	1203 ttanaagaaaacacgctgaacaacacacaccgcgcgcgtcccaaagaagaacacccg 1262
OY	345 tcgcagtgtaaccttaaacacatatcactttaaacacgtaagaagaagaagaagactctg 404

Tue Jan 23 10:54:41 2001

us-09-389-000-1.rng

Page 12

Db 1263 ctccagctccaaaacgagagatccagctcgtaaaacccaagcagaaaaaacacagctgac 1322

Qy 405 aaagtcagtcgaatatactatgcacttaaacctggaaacaaagatcaca 450

Db 1323 aaccaagctgaagaagactatgctcctgtagatcagaagaagaatatga 1368

Search completed: January 19, 2001, 03:27:27
Job time: 15374 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 22:07:46 ; Search time 1217.86 Seconds
(without alignments)
12244.374 Million cell updates/sec

Title: US-09-389-000-1

Perfect score: 2128
Sequence: 1 gaccggggggcgttggtggtc.....ccacaattatgaaaaaaa 2128

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: em_est40:*
- 42: em_estfun:*
- 43: em_esthum1:*

- 44: em_esthum2:*
- 45: em_esthum3:*
- 46: em_esthum4:*
- 47: em_esthum5:*
- 48: em_esthum6:*
- 49: em_esthum7:*
- 50: em_esthum8:*
- 51: em_esthum9:*
- 52: em_esthum10:*
- 53: em_esthum11:*
- 54: em_esthum12:*
- 55: em_esthum13:*
- 56: em_esthum14:*
- 57: em_esthum15:*
- 58: em_esthum16:*
- 59: em_esthum17:*
- 60: em_esthum18:*
- 61: em_esthum19:*
- 62: em_esthum20:*
- 63: em_esthum21:*
- 64: em_esthum22:*
- 65: em_esthum23:*
- 66: em_esthum24:*
- 67: em_esthum25:*
- 68: em_esthum26:*
- 69: em_esthum27:*
- 70: em_esthum28:*
- 71: em_esthum29:*
- 72: em_esthum30:*
- 73: em_esthum31:*
- 74: em_esthum32:*
- 75: em_esthum33:*
- 76: em_esthum34:*
- 77: em_esthum35:*
- 78: em_esthum36:*
- 79: em_esthum37:*
- 80: em_esthum38:*
- 81: em_esthum39:*
- 82: em_esthum40:*
- 83: em_esthum41:*
- 84: em_esthum42:*
- 85: em_esthum43:*
- 86: em_esthum44:*
- 87: gb_est41:*
- 88: gb_est42:*
- 89: gb_est43:*
- 90: gb_est44:*
- 91: gb_est45:*
- 92: gb_est46:*
- 93: gb_est47:*
- 94: gb_est48:*
- 95: gb_est49:*
- 96: gb_est50:*
- 97: gb_est51:*
- 98: gb_est52:*
- 99: gb_est53:*
- 100: gb_est54:*
- 101: gb_est55:*
- 102: gb_est56:*
- 103: gb_est57:*
- 104: gb_est58:*
- 105: gb_est59:*
- 106: gb_est60:*
- 107: gb_est61:*
- 108: gb_est62:*
- 109: gb_est63:*
- 110: gb_est64:*
- 111: gb_est65:*
- 112: em_esthum21:*
- 113: em_esthum22:*
- 114: em_esthum23:*
- 115: em_esthum24:*
- 116: em_esthum25:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.2	28.8	617	97	AM961069 EST373036
2	545.6	25.6	624	97	AM961078 EST373045
3	484.6	22.8	491	4	AA293855 zt61d10.r
4	398.4	18.7	490	7	AA417643 zV04a08.r
5	373.2	17.5	422	143	R13043 yf71h06.r1
6	307.2	14.4	316	5	AA339260 EST44524
7	280.4	13.2	492	7	AA398348 zt61d10.s
8	244.2	11.3	503	111	BE65112
9	228.4	10.7	441	15	AI023533
10	180	8.5	500	146	W03238
11	168.4	7.9	394	13	AA861485
12	163.6	7.7	440	7	AA402245 zt65h06.r
13	143.4	7.0	360	15	AI028190
14	143.4	6.7	512	23	AI636904
15	124.6	5.9	282	32	AV205711
16	122.2	5.7	235	98	BB015390
17	122	5.7	170	183	B42447
18	116.6	5.5	274	30	AV044749
19	114.2	5.4	715	156	AO375951
20	113.4	5.3	226	98	BB015549
21	95.6	4.3	452	150	AQ196491
22	90.8	4.2	354	184	B92720
23	58.4	2.7	121	15	AI047238
24	50	2.3	878	190	CNS0187R
25	50	2.3	1101	190	CNS0039G
26	49.8	2.3	647	38	AV674451
27	49	2.3	1080	190	CNS011Y
28	48.8	2.3	438	151	AQ267168
29	48	2.3	396	177	AZ277553
30	48	2.3	421	192	CNS03YNT
31	48	2.3	1101	190	CNS001EB
32	46.2	2.2	1101	190	CNS0182P
33	44.6	2.1	796	190	CNS0118D
34	44.2	2.1	747	190	CNS011RQ
35	44.2	2.1	1001	190	CNS0015H
36	43.8	2.1	1101	190	CNS00PFS
37	43.6	2.0	559	156	AQ371843
38	43.6	2.0	581	191	CNS01UPZ
39	43.6	2.0	768	191	CNS01VSE
40	43	2.0	799	190	CNS01ISA
41	43	2.0	1101	190	CNS003DO
42	43	2.0	1101	190	CNS006PA
43	43	2.0	1101	190	CNS01219
44	43	2.0	1309	108	BE420736
45	42.8	2.0	894	190	CNS018BG

ALIGNMENTS

RESULT 1
 LOCUS AM961069 617 bp mRNA
 DEFINITION EST373036 MAGF resequences, MAGF Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM961069
 VERSION AM961069.1 GI:8150648
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 617)
Hedge, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

TITLE

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL

unpublished (2000)

COMMENT

Contact: John Quackenbush

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johng@tigr.org

Plate: 152

Seq primer: Forward.

FEATURES

Location/Qualifiers

1..617

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGF"

/note="Vector: pBluescriptSKm"

BASE COUNT

211 a 116 c 127 g 163 t

ORIGIN

Query Match

Best Local Similarity 28.8%; Score 612.2; DB 97; Length 617;
Matches 614; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db

295 ccatggaactcttcagatagactatctctatctcgtctctatgaagatgcagatgga 354

Db

1 ccatggaactcttcagatagactatctctatctcgtctctatgaagatgcagatgga 60

Db

355 atccttaacacattacttaaacagctagaagaagaagaagaagaagaagaagaaga 414

Db

61 atccttaacacattacttaaacacacttagaagaaagaaagaaagaaagaaagaa 120

Db

415 gaataactatgacttaacacagctagaagaagaagaagaagaagaagaagaagaaga 474

Db

121 GAATACATGACATCTTAACCTGAGAACAGAAAGCTTACCGAAGATCAACATGA 180

Db

475 aagcgctacataccttagctgaagatgctcagggctcgtgttacaatcgaagttcctaaag 534

Db

181 ACGCGTACATACCTAGCTGAATGTCTCAGGGTCTGCTTACATCAAGTTCTAAAG 240

Db

535 gcaacaggtgatacactgacgctcagatgacagagatctagtgaaagcaaaataaac 594

Db

241 GCAACAGGTGATCAACCTGCTAGATGCAAGAGATCTAGTGAAGAACGCAAAATGAC 300

Db

595 atctatagctgagatgctcagctggtcctcagctgctgctgctgctgctgctgctg 654

Db

301 ATCTTATTAGTTGAGATGCTCAGTGGCTTACCTGCTGATACGTACAGAAATATT 360

Db

655 gcaacatagcagaagtcac 714

Db

361 GCAACATAGCAGAGATGACCATCACCATCAGTACGACAGAGAGAGAGAGAGAGAG 420

Db

715 gatattgcatatcacaatggtctctctggaagtgctcttcaactaagtcgagagag 774

Db

421 GATATTGTCATATTCATACATGCTCTTGAAGTGCTCTTCTTCACTAGTCCAGAGAG 480

Db

775 ctggaagcacaagttcaattagattggcaaaaagaataatacaatcactgctgttct 834

Db

481 CTGGAAGCACAAGTATTAATGATTGGCAAAAGAAAGAAATACATCTCACTGTTTGT 540

Db

835 ttataatccctgaaaatttcaaagtgatattcaagggcatggaatgataatgcttta 894

Db

541 TTTATTAATCCCTGAAAATTTTAAAGTGTATTCTCAGGGCATGAGATGATTTGCTTTA 600

Db

895 actgaacacatgacaat 911

Db 601 ACTGACCATGACAAAT 617

RESULT 2

LOCUS

AM961078 624 bp mRNA EST 01-JUN-2000

DEFINITION

EST373045 MAGE resequences, MAGF Homo sapiens cDNA, mRNA sequence.

ACCESSION

AM961078

VERSION

AM961078.1 GI:8150657

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 624)
Hedge, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

TITLE

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL

unpublished (2000)

COMMENT

Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johng@tigr.org
Plate: 152
Seq primer: Forward.

FEATURES

Location/Qualifiers

source

1..624

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGF"

/note="Vector: pBluescriptSKm"

BASE COUNT

209 a 117 c 128 g 170 t

ORIGIN

Query Match 25.6%; Score 545.6; DB 97; Length 624;
Best Local Similarity 96.3%; Pred. No. 1e-135;
Matches 591; Conservative 0; Mismatches 19; Indels 4; Gaps 3;

Db

295 ccatggaactcttcagatagactatctctatctcgtctctatgaagatgcagatgga 354

Db

1 ccatggaactcttcagatagactatctctatctcgtctctatgaagatgcagatgga 60

Db

355 atccttaacacattacttaaacagctagaagaagaagaagaagaagaagaagaagaaga 414

Db

61 atccttaacacattacttaaacacacttagaagaaagaaagaaagaaagaaagaaagaa 120

Db

415 gaataactatgacttaacacagctagaagaagaagaagaagaagaagaagaagaagaaga 474

Db

121 GAATACATGACATCTTAACCTGAGAACAGAAAGCTTACCGAAGATCAACATGA 180

Db

475 aagcgctacataccttagctgaagatgctcagggctcgtgttacaatcgaagttcctaaag 534

Db

181 ACGCGTACATACCTAGCTGAATGTCTCAGGGTCTGCTTACATCAAGTTCTAAAG 240

Db

535 gcaacaggtgatacactgacgctcagatgacagagatctagtgaaagcaaaataaac 594

Db

241 GCAACAGGTGATCAACCTGCTAGATGCAAGAGATCTAGTGAAGAACGCAAAATGAC 300

Db

595 atctatagctgagatgctcagctggtcctcagctgctgctgctgctgctgctgctg 654

Db

301 ATCTTATTAGTTGAGATGCTCAGTGGCTTACCTGCTGATACGTACAGAAATATT 360

Db

655 gcaacatagcagaagtcac 714

Db

361 GCAACATAGCAGAGATGACCATCACCATCAGTACGACAGAGAGAGAGAGAGAGAG 420

Db

715 gatattgcatatcacaatggtctctctggaagtgctcttcaactaagtcgagagag 774

Pb 421 GATGATTGCAATATCAACATGTTCTCTTGAAGTGCCCTTTTCACTAAGTGCCGAGAG 480
 QY 775 cgggaagccatcaagtaattag-attggcaaaagaaaataacacattcactg-cttg 832
 Db 481 CUGGAAGCCATCAAGTAATTAAGATTGGCAAAAGAAAATTCACAAATTCCTGTTTG 540
 QY 833 tttttaaacccttgaaaatttaaggtgtatcca--gggcacatgaaatgatttcg 890
 Db 541 TTTTAAATCCCTGAAAATTTTAAAGTTTCTTTTCAAGGCGCTGGAATGATTTTGT 600
 QY 891 tttaactgaaccac 904
 Db 601 TTTAACTTGAAACC 614
 RESULT 3
 AA293855 491 bp mRNA EST 12-AUG-1997
 LOCUS z61d10.r1 Soares_testis_NHT Homo sapiens cdna clone IMAGE:726835
 DEFINITION 5', mRNA sequence.
 ACCESSION AA293855
 VERSION AA293855.1 GI:1941833
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 491)
 Hillier, L., Allen, M., Bowles, L., Dubague, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyile,
 T., Waterston, R. and Wilson, R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 CONTACT: Wilson RK
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 833 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 446.
 FEATURES
 source
 1..491
 /organism="Homo sapiens"
 /db_xref="GDB:5923745"
 /db_xref="taxon:9606"
 /clone="IMAGE:726835"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pTZ19-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAGTGGAGCGGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTZ19 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 166 a 99 c 115 g 111 t.
 ORIGIN
 Query Match 22.8% Score 484.6; DB 4; Length 491;
 Best Job: 1 Similarity 99.2%; Pred.No. 2.4e-119;
 Matches 67; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 13 gttggggttcacgcctgctgactgctctctggtgctcctaattgtcttgcctc 72

Db 1 GTTGGGCTTCACCGCTCTGCGCTACTGCTTCTTGAGGGCCCTTAATGCTTGTC 60
 QY 73 taagtgctgaggggaaagacgcggaggctctctggtcctgcacataagaagaagaga 132
 Db 61 TAAGTGCTGCCGGAACACCGGGAGGCTCTGCGCTGACACTATGAGAGAGAGA 120
 QY 133 aactacaacttcgagcgtgtgagcaccacgcctgaaacacagctgtgtgaaagatc 192
 Db 121 AACTACAACTTCGACGCTGTGAGCACCAACCGCTGAACACAGAGTTGCTGAGAGATC 180
 QY 193 cgcgaagtagtgaatgggaaaacccgttatagacacacttgaatlaatgatgaact 252
 Db 181 CGCAAGAACTAGTATGATGAGAAAACCGTTATGACACAACTGAATTAATGATGAAT 240
 QY 253 agaaagcaaatgttlatctcaaggagaagtggaaaaatccatggaactcttcaga 312
 Db 241 AGAAAGCAAAATGTTTATCTCAAGAGAAAGTGGAAAATCCATGAAACTCTTCAGA 300
 QY 313 tagactatcttcattcgtgctatgaacgaatgcagtggaatccttaacacattact 372
 Db 301 TGACTATCTTCTATTCTGCTCTATGACGAATGCCAGTGAATCCTTAACACATTACT 360
 QY 373 taacacgctagaagaagaagaagactcttgaagtcaagtgaatatactatgacattaa 432
 Db 361 TAAACAGTAGAAGAGAAAGAGAGACTCTTGAAGTCAAGTGAATTAATGACTTAA 420
 QY 433 actggaacagaatcaaaagcttaccagaagaatcaaatgaagcgcgaactactaac 492
 Db 421 ACTGGAAACAGATCAAAAGCTTACCGAAGATCAACAAATGAAGCGCGTACACTAGC 480
 QY 493 tgaatgcttc 503
 Db 481 TGAATGCTTC 491
 RESULT 4
 AA417643 490 bp mRNA EST 02-MAR-1998
 LOCUS zv04a08.r1 Soares_NHMPu_s1 Homo sapiens cdna clone IMAGE:72630
 DEFINITION 5', mRNA sequence.
 ACCESSION AA417643
 VERSION AA417643.1 GI:2079462
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 490)
 Hillier, L., Allen, M., Bowles, L., Dubague, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Stepien, M., Tan, F., Theising, B.,
 White, Y., Wyile, T., Waterston, R. and Wilson, R.
 WashU-Merck human EST Project
 Unpublished (1997)
 CONTACT: Wilson RK
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 950 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 465.
 FEATURES
 source
 1..490
 /organism="Homo sapiens"
 /db_xref="GDB:5975465"
 /db_xref="taxon:9606"
 /clone="IMAGE:72630"
 /clone_lib="Soares_NHMPu_s1"

/tissue-type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT7T30-Pac
 (Pharmacia) with a modified polylinker; site_1: Not I;
 site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2Nbhm, pregnant uterus
 NBdhu, and fetal heart NBdhl9w) were mixed, and ss
 circles were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

[illegible]

	Tel: 314 286 1800	
	Fax: 314 286 1810	
	Email: estewatson.wustl.edu	
	Insert Size: 1034	
	High quality sequence stops: 343 Source: IMAGE Consortium, LNL	
	This clone is available royalty-free through LNL; contact the	
	IMAGE Consortium (infoimage.lnl.gov) for further information.	
	Insert Length: 1034 Std Error: 0.00	
	Seq primer: M13RP1	
FEATURES	High quality sequence stop: 343.	
source	Location/Qualifiers	
	1..422	
	/organism="Homo sapiens"	
	/db_xref="GDB:400152"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:27805"	
	/clone_lib="Scars Infant brain INIB"	
	/sex="female"	
	/dev_stage="73 days post natal"	
	/lab_host="DH10B (ampicillin resistant)"	
	/note="Organ: whole brain; Vector: Lambda Ba; Site_1: Not	
	I - Site_2: Hind III; 1st strand cDNA was primed with a Not	
	I - oligo(dT) primer [5';	
	AACGGAGAAATTCGCCGCCACCAGCAATTCTTTTCTTTTCTTTT 3'];	
	double-stranded cDNA was ligated to Hind III adaptors	
	(pharmacia), digested with Not I and directionally cloned	
	into the Not I and Hind III sites of the lambdaid BA vector	
	library went through one round of normalization. library	
	constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	138 a 80 c 95 g 109 t	
ORIGIN		
Query Match	17.5%; Score 373.2; DB 143; Length 422;	
Best Local Similarity	96.4%; Pred. No. 2e-89;	
Matches 403; Conservative	0; Mismatches 13; Indels 2; Gaps 2	
OY 891	ttaactgaaccactgacatggaaaaataagtaagtgtgtaaatactggacaacatg 950	
Db 1	TTTAACGTACCCACCTGCAGCAATGGAATAAATGAGTATGTGTAAATACTCGACAACATG 60	
OY 951	tcgcccaaacactgftaaggactgaaaagaacactggcctggaagaacttgattccccc 1010	
Db 61	TCCCTCAACACTGTGTAAGACTGAAAACGCAACTGGCCTGAGAACCTTGATGCCCT 120	
OY 1011	gcagaggtcctacacagcgacaacctgggatatttctactgactctactatttcctgctcta 1070	
Db 121	GCAGAGGTCTCACACACGAAACACCTGGGATATTCTCTACTGATCATTTGCTGCTGCTGA 180	
OY 1071	atctttaaggaatgycatggccttggaattaatgctctcgtgtgcaagttctggaaaaaa 1130	
Db 181	ATCTTTAAGGAATGGCAATGGCTTGAAATCTTCCTGTTGAGAGATTGGAGAAAAA 240	
OY 1131	caaaaagctctctctctctcatccaagaagaagaaactaagaaggaagaagatcaata 1190	
Db 241	CAAAAAGATCTCTCTCTCTTCATTCACGACAGAAAACCTAACAGGAAGATCAANA 300	
OY 1191	ttgctgtgagcaagctggtactctcttctgcglatatgtlaaaggagaaa-gaatgatgcg 1249	
Db 301	TTGCTGTGAGCAGCTGCTACTCTCTTCCGCTATTTAAAGGGAAGAAAGATGATGCGG 360	
OY 1250	cctcacgtctctga-ggacaagttgatattatggaatatatccggagagaatactctc 1306	
Db 361	CTTAGCTCTTGAGGCGCAACAGTTGATATATGTGGAAATATATACCGGGAGAAATATCTC 418	
RESULT 6	AA339260 316 bp mRNA EST 21-APR-1997	
LOCUS	AA339260 Fetal brain I Homo sapiens CDNA 5' end, mRNA sequence.	
DEFINITION	AA339260	
ACCESSION	AA339260.1 GI:1991678	
VERSION		
KEYWORDS	EST.	

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 316)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fullmer, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-ai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palanges, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Behnati, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dinke, D., Feng, D.-F., Ferrite, A., Fischer, P., Hastings, G.A., He, W.W., Hu, S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Mel, Y.F., Ming, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Pannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tcdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
FEATURES
SOURCE location/Qualifiers
1..316
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):141055"
/db_xref="taxon:9606"
/clone_lib="Fetal brain I"
/sex="female"
/dev_stage="fetus, 24 wks"
/note="Organ: brain, Vector: pluescript SK-, Site_1: EcoRI, Site_2: XhoI"
BASE COUNT 115 a 55 c 65 g 78 t 3 others
ORIGIN
Query Match 14.4%; Score 307.2; DB 5; Length 316;
Best Local Similarity 98.1%; Pred. No. 9.9e-72;
Matches 309; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 263 attgttatctcaagagaagtgaaaaaacatccatggaactcttcagatagactatct 322
Db 1 ATTTTTCACAGAGAAAGTGGAAAAATCCATGGAATTTTCAGATAGACTGTCT 60
QY 323 tctatctgtctatgaagcaatgcagtggaatcccttaaacacattactaaacagcta 382
Db 61 TCTATTCGTCTATGAGAGANTGCCAGTGAATCCTTTAAACATTACTTAACAGCTA 120
QY 383 gaagaagaagaagaactcttgaagaagtcgaataactatgcacttaactgtaacaa 442
Db 121 GAAGAAGAAGAAGAGACTCTTGAAGTCAAGTGAATATGATGCACTTAACTGGAACA 180
QY 443 gaataaagagcttaccagaagaatcaacaatgaagccgtaactatagctgaatgtct 502
Db 181 CAATCAAAAGCCTTACCAAGATCAACAATGAACGCCGCTACTAGCTAAATGTCT 240
QY 503 cagggttctgtttacatcaagttctaaagacacaggttgatcaactgctagatg 562
|||||

Db 241 CAGGTTCTGTTACATCAAGTTTCTAAAGCAGACAGTGAATCAATGCTAGATG 300
QY 563 caaggaatctagtg 577
Db 301 CAAGAGAACTAGTG 315
RESULT 7
AA398348/C
LOCUS
DEFINITION
AA398348 492 bp mRNA EST 12-AUG-1997
z61d10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:726835
3', mRNA sequence.
ACCESSION
AA398348
VERSION
AA398348.1 GI:2051457
KEYWORDS
EST.
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, T., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 833 Sd Error: 0.00
Seq primer: -41m13 fwd. EF from Amersham
High quality sequence stop: 372.
FEATURES
SOURCE location/Qualifiers
1..492
/organism="Homo sapiens"
/db_xref="GDB:5923745"
/db_xref="taxon:9606"
/clone="IMAGE:726835"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5'] TGTTACCAATCTGAAGGAGGAGGCGCCCAATTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 126 a 91 c 92 g 183 t
ORIGIN
Query Match 13.2%; Score 280.4; DB 7; Length 492;
Best Local Similarity 94.8%; Pred. No. 1.8e-64;
Matches 290; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 306 ctcaagatagactatcttctatctgtctatgaagaaatgcagtggaatcccttaaca 365
Db 492 CTTGAGATGACTATCTCTATTCGTCTATGAGAGATGCCAGTGAATCCTTAAACA 433
QY 366 catctactaaacagctagaagaagaagaagagactcttgaagaatgaatgaatctat 425
Db 432 CATTAAGCTTAACAGCTAAGAGAGAGAGAGAGCTTGAAGTCAAGTGAATTAATCTATG 373
QY 426 cacttaactggaacagaatcaagagcttaccagaagatcaacaatgaacgcgtacat 485
|||||


```

Db 372 CACTTAACAGGAGACAGAAATCAAGGCTTACAGAGATCAACAAATGAGCGCTACAT 313
QY 486 accatgctgaatgctcgaaggtctggtttacatcaagttctaaaggcaacagttg 545
Db 312 ACTTACTGAAATGCTCAGGAGGTTCTGTTTACATCAAGTTTCTAAAGGCAACAGGTGG 253
QY 546 atcaatgctgaagatgcaagaatctagtgaaacgcgcaaaatagacattctattagt 605
Db 252 ATCAACTGCTAGATGCAAGAGATCTAGTAAGAACTGAGATATTAATCAAGCTAGAC 193
QY 606 tggaga 611
Db 192 AGAAGA 187

RESULT 8
LOCUS BE665112 503 bp mRNA EST 08-SEP-2000
DEFINITION 153097 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE665112
VERSION BE665112.1 GI:10025321
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 503)
Smith,T.P.L., Casas,F., Stone,R.T., Heaton,M.P., Grose,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.W. and Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mismatch 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACACAT
BACKWARD: GTTTCACAGTACAGACG
Plate: 65 row: G column: 20
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
source 1..503
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."

BASE COUNT 159 a 108 c 101 g 135 t
ORIGIN
Query Match 11.5%; Score 244.2; DB 111; Length 503;
Best Local Similarity 78.6%; Pred. No. 9.6e-55;
Matches 359; Conservative 0; Mismatches 83; Indels 15; Gaps 5;

QY 1682 agaacattcaattcattcctcagccatgcccccgtct-caagcttcctcctgcgc 1740
Db 1 AGAACACTGACAGCGCTTCCCAAGAGCTGTACCAAGAGTCCAGTTCTTCCTCAGC 60
QY 1741 acgcgctctggtgtggcgcaagcgtgcactaccacactgtctcacaagtctt 1800

```

```

Db 61 ACTGCAATTCCTGTGTTGGCCAGCTGTGCACAGACACCCCAACTCTTCGCACATTTT 120
QY 1801 gggcgtatlaaaggcacatgtttgaatcaccctcaccactcactgagcgagttt 1860
Db 121 GGGCGTATTATATGACATATTGGAAT--GCACCTAGACACCATTTGGGTCGGCTTT 178
QY 1861 gacaatctggaaagctggaacaaagatgatttgaagctccaccaccaagactaata 1920
Db 179 GACAATACAGAGAGATGGAACAAAGATGATCTTAAAGCTACACCCAGACCTGATA 238
QY 1921 tcaaaagagtgacagtggttgcctctcgtataatcaccctcaagcttcgcagataga 1980
Db 239 TCAAAAGGTTGTCATTTGATGCTCTGTGATTAATGTCTAAAGCTTCGCGAGTAA 298
QY 1981 agaccagcagcaaaaagctgcccacac-----actgacatcattcaccatac 2030
Db 299 AGACAGGACAGCAGAGTGTGCTAGTACCCGATGCTTTTATTAATTAAGTTACATCA 358
QY 2031 acttgatcccgccagccagagagctacagaacaatgagctcagtgacccaactc- 2089
Db 359 ACTTGGATCTCC-CCAAACAGGCTGCTAAGAAACAAATGGCCTAGGTGACCTGACACT 417
QY 2090 tctttctcaaaaatattccacattatgaaaaa 2126
Db 418 TTTTCTCTCAAAACATTTCCACAGTTTATGCAAAA 454

RESULT 9
LOCUS A1023533 441 bp mRNA EST 27-AUG-1998
DEFINITION ov79d11.s1 Soares.testis_NHT Homo sapiens cDNA clone IMAGE:1643541
ACCESSION A1023533
VERSION A1023533.1 GI:3238577
KEYWORDS EST.
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 441)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 838 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 372.
Location/Qualifiers
source 1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1643541"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clonech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCCGCCCAATTTTATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library

```


BASE COUNT 85 a 87 c 84 g 138 t
 ORIGIN
 went through one round of normalization to Cct5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 7.9%; Score 168.4; DB 13; Length 394;
 Best Local Similarity 99.4%; Pred. No. 2.1e-34;
 Matches 169; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 415 gaattcttgccttaactgtaagaataaaggtcctcagaagtcacatga 474
 |||
 Db 394 GAAATCTATGCACTTAATGTAAGATCAAGGCTTACCAAGATCAACATGA 335
 Qy 475 acgcccatactactagctgaatgtctcagggtctgtgttcaatgaagttcctaag 534
 |||
 Db 334 ACCCCCTACATACCTAGCTGAATGTCCTCAGGGTCTGTGTTACATCAATTTCTTAAAG 275
 Qy 535 gcacacagtgatgacactgcctagatgcaagaatctagtgaacgc 584
 |||
 Db 274 GCACACAGTGATGATCACTGCTAGATGCAAGATCTGTGAAAAACC 225

RESULT 12
 AA402245 440 bp mRNA EST 12-AUG-1997
 LOCUS 265h06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727259
 DEFINITION 5', mRNA sequence.
 ACCESSION AA402245
 VERSION AA402245.1 GI:2056883
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Scheinberg, K., Stepien, M., Tan, F., Thaising, B., White, Y., Wyllie,
 T., Waterston, R. and Wilson, R.
 TITLE WashU-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert length: 561 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham.

FEATURES
 source
 location/Qualifiers
 1..440
 /organism="Homo sapiens"
 /db_xref="GDB:5924171"
 /db_xref="taxon:9606"
 /clone="IMAGE:727259"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5']
 TGTTCACATCTGAGTGGAGCGGCCCAATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cct5, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 94 a 110 c 135 g 101 t
 ORIGIN

Query Match 7.7%; Score 163.6; DB 7; Length 440;
 Best Local Similarity 97.6%; Pred. No. 4.2e-33;
 Matches 166; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 33 gccctactggtcttgcctgaatgtcttgcctcctaagtgctgagggagaaga 92
 |||
 Db 9 GCTATCATGGCTTCTGGGCGCCCTTAATGCTGTGCTCTAAGGTGAGCGGAAGA 68
 Qy 93 cgggggaaggtcttgcctgacactatgaagaaggaataacttgacgttt 152
 |||
 Db 69 CGCGGAGGCTCTTGCCCTGACACTATGAAGGAAGAACTACAACTTGACCGGT 128
 Qy 153 gaaccacacccgctaaacacagtgctggaagaagccgcaagaagt 202
 |||
 Db 129 GAGCACCAACCGCTTAACACAGAGTGTGGAAGAAGTCCCAACAACT 178

RESULT 13
 A1028190 360 bp mRNA EST 27-AUG-1998
 LOCUS ov90c10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1644594
 DEFINITION 3', mRNA sequence.
 ACCESSION A1028190
 VERSION A1028190.1 GI:3245499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bbrp/image/image.html
 Insert length: 874 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 320.

FEATURES
 source
 location/Qualifiers
 1..360
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1644594"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5']
 TGTTCACATCTGAGTGGAGCGGCCCAATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cct5, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 92 a 67 c 69 g 132 t
 ORIGIN

Query Match 7.0%; Score 148; DB 15; Length 360;
 Best Local Similarity 91.3%; Pred. No. 6.2e-29;

Matches 157; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 440 caagaatcaaaagcttaccagaagaatcaacaataaagccgttacatctgctgaatg 499
 |||||
 Db 360 CAAGAATCAAAAGGCTTACCAAGAAATCAACAATGAAACCCCTACATACCTGCTGAATG 301
 |||||

QY 500 tctcaaggttcgtgttccatcaagttctcaaaaggaacaggttgatcaactgcctag 559
 |||||
 Db 300 TCTAGGGTTGTGGTTTACATCAAGTTTCTAAAAGGCAACAGGTGATCAACTGCTTAG 241
 |||||

QY 560 atgcaagaagaatctagtagaaacgcaaaataagacatcttagttgaga 611
 |||||
 Db 240 ATGCAAGAGATCTAGTGAATACTGGAAGATATATATCCAGCTAAGCAGAGA 189
 |||||

RESULT 14
 A1636904 512 bp mRNA EST 26-APR-1999
 LOCUS ts90006.x1 NCI_CGAP_G6 Homo sapiens cDNA clone IMAGE:2238539 3',
 DEFINITION mRNA sequence.
 ACCESSION A1636904
 VERSION A1636904.1 GI:4688234
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 512)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 460.
 Location/Qualifiers
 1. 512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2238539"
 /clone_lib="NCI_CGAP_G6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: p773D-PAC (Pharmacia) with a modified
 polylinker. Plasmid DNA from the normalized library
 NCI_CGAP_G64 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (clonoids 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo.

BASE COUNT 113 a 111 c 111 g 177 t

ORIGIN

Query Match 6.7%; Score 143.4; DB 23; Length 512;
 Best Local Similarity 99.3%; Pred. No. 1.2e-27;
 Matches 144; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 442 agaatcaaaagcttaccagaagaatcaacaataaagccgttacatctgctgaatg 501
 |||||

Db 512 AGAATCAAAAGCTTACCAGAAGATCAACAATGAACCGCTACATACCTGAAATGTC 453
 |||||

QY 502 tcaaggttcgtgttaccataaagttctaaaggcaacaggttgatcaactgcttagat 561
 |||||

Db 452 TCAGGGTTCTGGTTTACATCAAGTTTCTAAAAGGCAACAGGTGATCAACTGCTTAGAT 393
 |||||

QY 562 gcaagaatctagtagaaacgcaa 586
 |||||

Db 392 GCAAGAGATCTAGTGAATAACGCA 368
 |||||

RESULT 15
 AV205711 282 bp mRNA EST 30-OCT-1999
 LOCUS AV205711
 DEFINITION cDNA clone 1700081L05 3', mRNA sequence.
 ACCESSION AV205711
 VERSION AV205711.1 GI:6146564
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 282)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
 Iishi, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
 C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
 Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
 Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
 Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al. 1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 URL: http://genome.rtc.riken.go.jp/
 Sasaki, N., Izawa, M., Watanabe, M., Okazaki, Y., and Hayashizaki,
 Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki,
 Y.

Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.
 Location/Qualifiers
 1. 282
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1700081L05"
 /clone_lib="RIKEN full-length enriched, adult male testis"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: BamHI; cDNA library was

THIS PAGE BLANK (USPTO)